

Hinweise zur Aufbereitung quantitativer Daten

Ebel, Thomas; Trixa, Jessica

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Ebel, T., & Trixa, J. (2015). *Hinweise zur Aufbereitung quantitativer Daten*. (GESIS Papers, 2015/09). Köln: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.21241/ssoar.43223>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC Lizenz (Namensnennung-Nicht-kommerziell) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-nc/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC Licence (Attribution-NonCommercial). For more information see: <https://creativecommons.org/licenses/by-nc/4.0>

Hinweise zur Aufbereitung quantitativer Daten

Thomas Ebel und Jessica Trixa

GESIS Papers 2015|09

Hinweise zur Aufbereitung quantitativer Daten

Thomas Ebel und Jessica Trixa

GESIS Papers

GESIS – Leibniz-Institut für Sozialwissenschaften
Datenarchiv für Sozialwissenschaften
Unter Sachsenhausen 6-8
50667 Köln
Telefon: (0221) 476 94 - 411 / - 495
Telefax: (0221) 476 94 - 199
E-Mail: thomas.ebel@gesis.org / jessica.trixa@gesis.org

ISSN: 2364-3781 (Online)
Herausgeber,
Druck und Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln

Inhalt

1	Hinweise zur Aufbereitung quantitativer Daten	5
1.1	Übersicht der zu übermittelnden Daten und Dokumente	5
1.2	Checkliste der notwendigen Arbeitsschritte für die Bereitstellung quantitativer Daten	5
1.3	Erläuterungen zur Aufbereitung quantitativer Daten	7
1.3.1	Variablen- und Wertebenennung	7
1.3.2	Variablen- und Wertelabels	8
1.3.3	Variablenwerte	8
1.3.4	Fehlende Werte	9
1.3.5	Dokumentationsmaterialien	9
1.3.6	Datenschutz und rechtliche Aspekte	9
1.3.7	Plausibilitäts- und Konsistenzprüfungen	10
1.3.8	Dateinamen	10
1.3.9	Dateiformat	11
1.4	Literatur	11
1.5	Fußnoten	12

1 Hinweise zur Aufbereitung quantitativer Daten¹

Um die Nachvollziehbarkeit von Forschungsdaten für SekundärnutzerInnen, d. h. für nicht an der Erhebung beteiligte ForscherInnen sicherzustellen, müssen diese entsprechend aufbereitet und die dazugehörigen Dokumentationsmaterialien in konsistenter und nachvollziehbarer Weise zur Verfügung gestellt werden. Hierbei sollten einige Mindestanforderungen beachtet werden. Vor diesem Hintergrund liefert dieses Dokument

1. eine Übersicht über die einzureichenden Daten und Dokumentationen,
2. eine Checkliste mit den notwendigen Arbeitsschritten für die Bereitstellung quantitativer Daten und
3. detaillierte Erläuterungen zur Aufbereitung quantitativer Daten.

1.1 Übersicht der zu übermittelnden Daten und Dokumente

- Aufbereiteter und fehlergeprüfter/-korrigierter Datensatz in akzeptablem Format (bspw. R, SAS, SPSS oder STATA) [\[1\]](#)
- Instrumente der Datenerhebung, wie Fragebögen, Listen- und Kartensätze, Testinstrumente
- Ggf. Dokumentation der Variablen (sog. Codebuch, Datenhandbuch oder Skalenhandbuch), inkl. Benennung aller Variablen, Wortlaut der Items im Fragebogen, Werteausprägungen und Codierung sowie Benennung fehlender Werte [\[2\]](#)
- Ggf. Methoden-/ Feldbericht [\[3\]](#): Angaben zu Untersuchungsdesign und zur Feldphase, insbesondere mit Angaben zur Stichprobe [\[4\]](#)
- Ggf. weitere Dokumente zur Beschreibung der Studie (Syntaxen, Angaben zur Einhaltung datenschutzrechtlicher Vorgaben, Maßnahmen zur Anonymisierung, Angaben zu durchgeführten Plausibilitäts-, Konsistenz- und Fehlerkontrollen, Zwischen- und Abschlussberichte)

1.2 Checkliste der notwendigen Arbeitsschritte für die Bereitstellung quantitativer Daten

Im Folgenden finden Sie eine Übersicht der wesentlichen Arbeitsschritte, die vorgenommen werden müssen, damit Ihre Daten nachvollziehbar sind. Detaillierte Hinweise zu jedem Punkt erfolgen im anschließenden Abschnitt.

¹ Wir danken herzlich Alexia Meyermann und Maike Porzelt (DIPF) sowie Claudia Neuendorf, Lisa Pegelow und Poldi Kuhl (IQB) für ihre Unterstützung. Dieses Dokument wurde im Rahmen des Verbunds Forschungsdaten Bildung erstellt. Der Verbund wird gefördert durch das Bundesministerium für Bildung und Forschung.

Dimension	Aspekte
Variablen- und Wertebenennung	<ul style="list-style-type: none"> ○ Variablen und Werte sind nach einem konsistenten Schema verständlich und eindeutig bezeichnet (z. B. V1 bis V100). ○ Alle Variablen sind über ihren Namen und/oder ihre Labels den jeweiligen Items im Fragebogen bzw. den Konstrukten im Skalenhandbuch zuordenbar. ○ Die Namen sind möglichst kurz gewählt. ○ Sonderzeichen, Umlaute oder Leerzeichen wurden nicht verwendet.
Variablen- und Wertelabels	<ul style="list-style-type: none"> ○ Idealerweise: Variablenlabels wurden für alle Variablen verwendet. ○ Idealerweise: Alle Werte sind gelabelt. ○ Variablen- und Wertelabels sind eine kurze und möglichst aussagekräftige Beschreibung der Variableninhalte bzw. -ausprägungen. ○ Variablen- und Wertelabels enthalten keine Sonderzeichen, Umlaute und Leerzeichen.
Variablenwerte	<ul style="list-style-type: none"> ○ Alle Variablenausprägungen sind jeweils einem eindeutigen numerischen Wert zugewiesen. ○ Die zugewiesenen numerischen Werte folgen, soweit möglich, einem einheitlichen Schema. ○ Offene Antwortmöglichkeiten sind ggf. codiert, d. h. mit einem numerischen Wert versehen worden. ○ Offene Antwortmöglichkeiten sind auf datenschutzrechtliche Probleme untersucht worden.
Fehlende Werte	<ul style="list-style-type: none"> ○ Fehlende Werte sind definiert. ○ Fehlende Werte sind als solche durch ein Label gekennzeichnet. ○ Existieren verschiedene Arten von fehlenden Werten, sind diese möglichst differenziert festgehalten.
Dokumentationsmaterialien	<ul style="list-style-type: none"> ○ Instrumente der Datenerhebung, bspw. Fragebögen, werden eingereicht. ○ Ein Codebuch oder Skalenhandbuch enthält alle im Datensatz auftretenden Variablen. ○ Zu jeder Variable ist deutlich, welcher Frage im Fragebogen sie entspricht und wie die Antwortalternativen kodiert wurden. ○ Falls Variablen erstellt wurden (z. B. Skalenbildung, Index, abgeleitete Variablen) ist dies im Code-/Skalenhandbuch nachvollziehbar beschrieben. ○ Angaben zum Untersuchungsdesign und der Feldphase sind in einem Methodenhandbuch/Feldbericht festgehalten.

Dimension	Aspekte
Datenschutz und rechtliche Aspekte	<ul style="list-style-type: none"> ○ Eigennamen von Personen wurden gelöscht und durch nicht sprechende Identifikatoren ersetzt, z. B. id1 bis id198 (formale Anonymisierung). ○ Weitere personenbezogene Daten (u. a. Eigennamen von Orten und Organisationen) liegen nicht vor oder die Studie wurde anonymisiert (formale Anonymisierung). ○ Regionale, berufliche und ähnliche Angaben sind nicht so kleinteilig, dass die Informationen die Identifizierung der TeilnehmerInnen ermöglichen (andernfalls sind Hinweise an das Archiv erfolgt). ○ Es liegen keine urheberrechtlichen oder vertraglichen Hindernisse vor, die einer Archivierung oder Datenweitergabe im Wege stehen.
Plausibilitäts- und Konsistenzprüfungen	<ul style="list-style-type: none"> ○ Die Filterführung ist korrekt. ○ Die Variablenausprägungen sind plausibel. ○ Es sind keine wild codes (Werte außerhalb des zulässigen Wertebereichs) vorhanden.
Dateinamen	<ul style="list-style-type: none"> ○ Dateinamen enthalten keine Sonderzeichen, Umlaute oder Leerzeichen. ○ Dateinamen sind möglichst kurz gewählt. ○ Dateien sind so bezeichnet, dass aus dem Namen Rückschlüsse auf Studie, Daten-/Materialtyp und ggf. Versionsnummer gezogen werden können.
Datenformat	<ul style="list-style-type: none"> ○ Die Dateien liegen in einem empfohlenen oder zumindest akzeptablen Format vor.

Abbildung 1: Checkliste der notwendigen Arbeitsschritte

1.3 Erläuterungen zur Aufbereitung quantitativer Daten

1.3.1 Variablen- und Wertebenennung

Die Variablen sollten möglichst verständlich und eindeutig benannt sein. Klare Konventionen vereinfachen zugleich die Erschließung und Nachnutzung der Daten. Mehrere Möglichkeiten sind zu unterscheiden.

1. Die Variable wird nach der Fragenummer benannt (z. B. F1 bis Fn). Somit wird ein direkter Bezug der Variable zur Originalfrage hergestellt und deren Reihenfolge im Fragebogen abgebildet.
2. Eine weitere übliche Art der Benennung ist die aufsteigende Nummerierung mit einem voranstehenden Buchstaben, z. B. „V“ für Variable. Auf diese Weise wird eine einfache Reihenfolge der Variablen im Datensatz abgebildet, allerdings können die Variablen nicht nach Inhalt bzw. Typ unterschieden werden. Daher werden oftmals zusätzliche inhaltliche Kürzel als weiterer Namensbestandteil genutzt.
3. Inhaltliche Kürzel: Diese sogenannten mnemotechnischen Variablennamen bieten sich vor allem bei Längsschnittanalysen an, wenn Fragemodule wiederholt eingesetzt werden, beispielsweise "B_EKOM" für das Einkommen des Befragten („B“ für Befragter, „EKOM“ für Einkommen).

men). Existieren thematisch zusammenhängende Variablenblöcke, z. B. bei Ländervergleichen, bietet sich eine thematische oder strukturelle Kennzeichnung über die Verwendung von Präfixen, Wortstämmen, etc. an. Beispielsweise heißt die länderspezifische Variable zur Parteienneigung des International Social Survey Programme für Österreich „AT_PRTY“ und verfügt zusätzlich über das Label „Country specific party affiliation: Austria“, die entsprechende Variable für Belgien heißt BE_PRTY, usw. (ISSP 2010a, [doi:10.4232/1.11418](https://doi.org/10.4232/1.11418)) (Jensen 2012: 27f.).

Generierte Variablen, die keiner Frage im Fragebogen entsprechen, sollten in jedem Fall zudem in einem zusätzlichen Dokument (dem sogenannten Code- oder Skalenhandbuch) beschrieben und entsprechend ausreichend gelabelt werden. Bei der Codierung von Ländern, Berufen, Bildungsangaben usw. ist es empfehlenswert, national oder international akzeptierte Klassifikationssysteme zu nutzen, wie z. B. CASMIN, Comparative Analysis of Social Mobility in Industrial Nations, oder ISCED, International Standard Classification of Education.

1.3.2 Variablen- und Wertelabels

Unabhängig von der Art und Weise der Variablenbenennung, aber insbesondere wenn keine sprechenden Variablennamen verwendet wurden, sollten zusätzlich erläuternde Variablenlabels genutzt werden. Durch die „Etikettierung“ von Variablen durch Labels sollen die Inhalte der Variablen durch eine kurze und möglichst aussagekräftige Beschreibung angegeben werden. So können die Variablen auch ohne Hinzuziehen von Fragebogen oder Codebuch verstanden werden, z. B. verfügt die Variable V24 des ISSP 2008 Religionsmoduls über das Variablenlabel „Q11c Religions bring conflict“ („Q“ steht für Question, „11c“ steht für die Nummer der Frage im entsprechenden Fragebogen).

In die Labels können u. a. die Fragenummer aus dem Fragebogen oder Hinweise zu Art oder Besonderheiten einer Variable aufgenommen werden, z. B., ob sie neu gebildet oder recodiert wurde (vgl. Jensen 2012).

Zudem werden im Idealfall alle Werte mit kurzen, dabei möglichst aussagekräftigen Labels versehen. [5] Teilweise ist dies nicht möglich, z. B. bei kontinuierlichen Variablen wie Einkommen und Alter. Bei dokumentierten Skalen wird es oftmals als ausreichend angesehen, nur die Endpunkte der Skala zu labeln (z. B. 1=sehr gut, 7=sehr schlecht).

1.3.3 Variablenwerte

Für statistische Auswertungen müssen den Antwortkategorien der Fragen numerische Werte zugewiesen werden. Die numerischen Codes müssen alle möglichen Antworten und fehlenden Werte der Frage umfassen, sich gegenseitig ausschließen und eindeutig sein (Jensen 2012: 29). Außerdem sollte die Codierung, soweit möglich, einheitlichen Schemata folgen. [6]

Offene Fragen (sog. String-Variablen) sind Fragen mit offener Antwortmöglichkeit. Sie bilden bezüglich der Codierung unter Umständen eine Ausnahme, da sie sich teilweise nicht in einfacher und sinnvoller Weise in numerische Codes umwandeln lassen. Sie werden dann u. U. nicht umcodiert, um keinen Informationsverlust zu verursachen, sondern in ursprünglicher Form belassen, dann aber ohne statistische/quantitative Auswertungsmöglichkeiten, nur teilweise umcodiert oder aber vollständig aus dem Datensatz entfernt. Zudem bergen sie das Risiko, dass sie Informationen enthalten, die zu einer Identifizierung der TeilnehmerInnen führen könnten. Offene Angaben sind daher sorgfältig auf datenschutzrechtliche Probleme zu untersuchen.

1.3.4 Fehlende Werte

Anschließend sollten fehlende Werte in den Variablen überprüft, definiert und dokumentiert werden. Ihnen werden in den entsprechenden Variablen spezielle Codes zugewiesen. Die Empfehlung hierzu ist entweder die Verwendung von numerischen Codes, die sich außerhalb des jeweiligen gültigen Wertebereiches der Variable befinden [7] oder aber negativer Werte (Jensen 2012: 31). Außerdem müssen auch fehlende Werte durch angemessene Wertelabels inhaltlich dokumentiert werden. Im Sinne einer strukturierten Datenkontrolle und einer späteren Datenanalyse sollten alle Missing Values möglichst differenziert erfasst werden. System Missings, d. h. nicht spezifizierte fehlende Werte, können auf diese Weise ausgeschlossen werden (Jensen 2012: 30). Beispiele für vers. Arten fehlender Werte sind „Keine Angabe“, „Weiß nicht“, „Trifft nicht zu“ (Filterführung), „Split“ (Splits der Stichprobe) und „Angabe verweigert“.

1.3.5 Dokumentationsmaterialien

Material, das bei der Datenerhebung genutzt wurde, bspw. Fragebögen, Leitfäden, etc., unterstützt die Interpretierbarkeit der Daten und sollte archiviert werden. Um die Nachnutzbarkeit eines Datensatzes zu maximieren, empfiehlt es sich darüber hinaus, zusätzliches Dokumentationsmaterial zu erstellen. So ist für Forscher eindeutig nachvollziehbar, wie die Variablen im Datensatz entstanden sind und welche Bedeutung verschiedene Ausprägungen tragen. Außerdem können hier Informationen untergebracht werden, die im Datensatz selbst keinen Platz finden.

Jede Variable, die im Datensatz enthalten ist, sollte im Codebuch bzw. Skalenhandbuch zu finden sein. Dabei sollten mindestens die Variablennamen, ggf. die Zuordnung zu den entsprechenden Fragen im Fragebogen, die möglichen Ausprägungen und die Codierung der Antworten im Datensatz angegeben sein. Häufig findet man eine Übersicht über die absoluten und relativen Häufigkeiten der Besetzung der Antwortkategorien, die Anzahl fehlender Werte, teilweise auch Item-Kennwerte wie Mittelwert und Standardabweichung. Falls Skalen gebildet wurden, wird im Skalenhandbuch festgehalten, welche Items in einer Skala zusammengefasst wurden und wie das entsprechende Konstrukt bezeichnet wird. Angegeben werden weiterhin die Methode der Skalenbildung sowie Skalenkennwerte (Mittelwert, Standardabweichung, Reliabilität). Weiterhin besteht im Skalenhandbuch die Möglichkeit, die Quellen der verwendeten Items und Skalen anzugeben.

Angaben zum Untersuchungsdesign (bzw. der Feldphase), insbesondere zur Stichprobe, Stichprobenziehung und ggf. Gewichtung, sollten in einem Methodenhandbuch/Feldbericht festgehalten werden. Dies erlaubt Nachnutzern die methodischen Aspekte der Studie einzuordnen und zu bewerten, bspw. ggf. die Repräsentativität der Stichprobe für die Grundgesamtheit.

Beispiele für ausführliche Datensatzdokumentationen finden sich in der Link-Leiste am rechten Bildschirmrand.

1.3.6 Datenschutz und rechtliche Aspekte

Zunächst sollte geklärt sein, unter welchen rechtlichen Rahmenbedingungen die Daten erfasst wurden. Die Forschung sollte auf der Grundlage der informierten Einwilligung der TeilnehmerInnen in die Studienteilnahme durchgeführt worden sein. [8] Außerdem wurden idealerweise Einwilligungen in die Archivierung und Datenweitergabe eingeholt. [9]

Der Umgang mit den erhobenen Forschungsdaten erfordert eine besondere Beachtung der rechtlichen Aspekte zum Persönlichkeitsschutz der Befragten. Besonders sensitiv sind hierbei personenbezogene Daten. Die Mindestvoraussetzung für die Übergabe von Forschungsdaten an das Datenarchiv von GESIS ist die formale Anonymisierung. Formales Anonymisieren umfasst das Entfernen aller direkten

Identifikatoren (Namen, Anschrift, Kontaktdaten, Registernummern etc.), z. B. durch Codierung von Befragten (Befragten-ID). Sollen Befragte zu weiteren Befragungen (bspw. im Rahmen eines Panels) eingeladen werden, müssen ihre Angaben getrennt von ihren Kontaktinformationen gespeichert und verarbeitet werden. Dazu werden Betroffene durch Pseudonymisierung verschlüsselt codiert. Sofern die personenbezogenen Kontaktdaten nicht für weitere Forschungsvorhaben benötigt werden, sollten sie so früh wie möglich gelöscht werden.

Für die Wahrung der Anonymität der Befragten ist es entscheidend, dass kleinteilige Informationen nicht veröffentlicht werden. Hierzu zählen insbesondere detaillierte Angaben zu Beruf und geographischen Regionen (Wohnort, Arbeitsort, etc.). Maßnahmen, die im Sinne einer Anonymisierung durchgeführt werden, sind beispielsweise die Vergrößerung von Antwortkategorien (z. B. durch Bildung von Einkommens- oder Altersgruppen) und Orts- bzw. Regionalangaben sowie die Kategorisierung von Berufsangaben, z. B. durch standardisierte Klassifikationsschemata wie die ISCO-Codierung (Jensen 2012: 66f.). [\[10\]](#)

Es sollten keinerlei (urheber-)rechtliche Hindernisse vor der Übergabe bestehen, wie etwa vertragliche Verpflichtungen gegenüber einem Dateneigentümer oder Geldgeber oder (rechtliche) Beschränkungen anderer Art, wie beispielsweise die Nutzung geschützter Skalen und Instrumente, die nicht ohne Erlaubnis Dritter veröffentlicht werden dürfen.

1.3.7 Plausibilitäts- und Konsistenzprüfungen

Plausibilitäts- und Konsistenzprüfungen sind nach Abschluss der Datenerhebung zur Aufbereitung und Bereinigung der Rohdaten sowie nach jeder größeren Veränderung des Datensatzes durchzuführen, um sicherzustellen, dass die Daten korrekt erfasst und durch nachfolgende Arbeitsschritte nicht in unzulässiger Weise verändert wurden. Dabei ist insbesondere auf eine korrekte Filterführung, die Plausibilität der Häufigkeiten der Variablenausprägungen und das Vorhandensein von wild codes, das sind Werte außerhalb des gültigen Wertebereichs, zu achten. [\[11\]](#)

1.3.8 Dateinamen

Ein Dateiname sollte möglichst kurz gewählt werden, da bei langen Datei- und Ordnernamen unter Umständen Probleme bei automatisierten Backup-Abläufen auftreten können. Außerdem sollte auf Sonderzeichen (mit Ausnahmen von Unter- und Bindestrichen), Umlaute und Leerzeichen verzichtet werden. Der Name sollte sich, je nach spezifischem Kontext, aus den folgenden Bestandteilen zusammensetzen:

1. ID oder Studiennummer: zur Zuordnung zur Studie
2. Kürzel für Daten- bzw. Materialtyp: Interview, Video, Fragebogen etc.
3. Laufende Nummer von Datentypen: 001 ff.
4. Seriennummer: verschiedene Dateien je Datentyp: z. B. besteht ein Video aus mehreren Dateien a, b, c ff.
5. Versionsnummer: falls Änderungen durchgeführt werden und dokumentiert werden sollen, zum Beispiel durch Anonymisierung

Beispiel: [ID oder Studiennummer]_[Kürzel für Daten- bzw. Materialtyp]_[laufende Nummer]_[ggf. Seriennummer]_[ggf. Version]

1.3.9 Dateiformat

Die Wahl des Formats hängt unter anderem vom Datentyp ab. Dabei gilt es zu bedenken, dass digitale Formate sich mit der Zeit ändern, möglicherweise obsolet werden und dann im schlimmsten Fall nicht mehr les- und nutzbar sind. Außerdem sind nicht alle Formate in gleichem Maße für die Bereitstellung von Daten und Dokumentation geeignet. ForscherInnen sollten sich um einheitliche, in der Fachdisziplin als Standard für den entsprechenden Datentyp geltende, möglichst offene oder zumindest portierbare Formate bemühen. Weitere Informationen zu empfohlenen Dateiformaten finden Sie unter www.gesis.org/unser-angebot/archivieren-und-registrieren/datenarchivierung/vorbereitung-datenuebergabe.

Besondere Vorsicht walten lassen sollten Sie bei der Konvertierung zwischen Formaten, da es bei diesem Schritt zu Informationsverlusten durch Zeichenbeschränkungen, fehlender Beachtung von Groß- und Kleinschreibung, nicht gegebener Darstellbarkeit von Sonderzeichen etc. kommen kann.

1.4 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: www.gesetze-im-internet.de/bdsg_1990, abgerufen am 26.11.2014.
- GESIS – Leibniz-Institut für Sozialwissenschaften, Datenarchiv. Online-Dokument: www.gesis.org/unser-angebot/archivieren-und-registrieren/datenarchivierung/vorbereitung-datenuebergabe, abgerufen am 12.12.2015.
- Häder, Michael, 2009: Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. RatSWD – Working Paper No. 90. Online-Dokument: www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf, abgerufen am: 19.01.2015.
- Jensen, Uwe, 2012: Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten. GESIS-Technical Reports 2012|07. Online-Dokument: www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf, abgerufen am 20.11.2014.
- Katharina Kinder-Kurlanda & Oliver Watteler, 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.
- Metschke, Rainer, Wellbrock Rita, 2002: Datenschutz in Wissenschaft und Forschung. Materialien zum Datenschutz Nr. 28., 3. Aufl. Berlin, 2002: www.datenschutz-berlin.de/attachments/47/Materialien28.pdf?1166527077, abgerufen am 26.11.2014.
- Quandt, Markus, Mauer, Reiner, 2012: Sozialwissenschaften. In: Neuroth, Heike; Strathmann, Stefan; Oswald, Achim; Scheffel, Regine; Klump, Jens; Ludwig, Jens (Hrsg.), Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme. 2012, Göttingen, S. 61-81: nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor_lza_forschungsdaten_bestandsaufnahme.pdf, abgerufen am 26.11.2014.
- Schnell, Rainer, 2012: Survey-Interview. Methoden standardisierter Befragungen. VS Verlag: Wiesbaden.
- Watteler, Oliver, 2010: Erstellung von Methodenberichten für die Archivierung von Forschungsdaten. GESIS. Online-Dokument: www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Aufbau_Methodenbericht_v1_2010-07.pdf, abgerufen am: 21.01.2015.

1.5 Fußnoten

[1] Siehe auch: Empfehlungen zu Dateiformaten (www.gesis.org/unser-angebot/archivieren-und-registrieren/datenarchivierung/vorbereitung-datenuebergabe, Zugriff am 08.05.2015)

[Zurück zum Text](#)

[2] Beispiele für entsprechende Dokumentationsmaterialien finden Sie hier: EVS (2011): European Values Study 2008: Integrated Dataset (EVS 2008). GESIS Datenarchiv, Köln. ZA4800 Datenfile Version 3.0.0, [doi:10.4232/1.11004](https://doi.org/10.4232/1.11004). Falls kein Codebuch/Datenhandbuch vorhanden ist, achten Sie bitte darauf, dass die Variablen im Datensatz den Fragen im Fragebogen zuzuordnen sind. Entweder über die Angabe der Variablennamen im Fragebogen selbst, über die Benennung der Variablennamen entsprechend der Fragenummerierung im Fragebogen oder über die Nennung der Fragenummer im Variablenlabel.

[Zurück zum Text](#)

[3] Hinweise zum Erstellen eines Methodenberichts finden sich bspw. bei Schnell (2012: 415ff.) oder bei Watteler 2010. Ein Beispiel für einen Methodenbericht (Codehandbuch) findet sich hier: Eberhard, Verena; Beicht, Ursula; Krewerth, Andreas; Ulrich, Joachim Gerd (2014): BIBB-Übergangsstudie 2011. GESIS Datenarchiv, Köln. ZA5916 Datenfile Version 1.0.0, [doi:10.7803/202.11.1.1.10](https://doi.org/10.7803/202.11.1.1.10).

[Zurück zum Text](#)

[4] Empfohlen wird die Angabe der Ausschöpfungsquote nach AAPOR Standard (siehe www.aapor.org/AAPORKentico/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx, Zugriff am 03.02.2015).

[Zurück zum Text](#)

[5] Bezüglich der Variable Geschlecht werden beispielsweise die Werte 0 und 1 als „Mann“ bzw. „Frau“ gelabelt.

[Zurück zum Text](#)

[6] Beispiel: Ja/Nein-Fragen immer Ja=1, Nein=0 codieren. Mehr Informationen zum Thema einheitliche Werteschemata finden Sie bei Jensen 2012: Abschnitt 2.1.5.

[Zurück zum Text](#)

[7] Beispiel in Anlehnung an Jensen (2012: 31): Geht der gültige Wertebereich bis zur Zahl 5, werden fehlende Werte z. B. als 7=verweigert, 8=weiß nicht und 9=keine Angabe codiert. Umfasst der gültige Wertebereich auch zweistellige Zahlen, wählt man 97, 98 und 99 (sofern diese nicht zum gültigen Wertebereich zählen) usw.

[Zurück zum Text](#)

[8] Informationen zur informierten Einwilligung bieten Metschke/Wellbrock 2002, Häder (2009: 16ff) und Jensen (2012: 14). Vorlagen für Einwilligungserklärungen in Studienteilnahme finden Sie ebenfalls bei Metschke/Wellbrock 2002 (Anlage 1).

[Zurück zum Text](#)

[9] Dabei handelt es sich um den Idealfall. Werden Forschungsdaten faktisch anonymisiert, sind keine Einwilligungen in Archivierung und Datenweitergabe erforderlich (Jensen 2012: 67).

[Zurück zum Text](#)

[10] Weiterführende Hinweise zu den Themen Datenschutz und Anonymisierungsstrategien erhalten Sie in Kinder-Kurlanda/Watteler 2015.

[Zurück zum Text](#)

[11] Weitere Informationen zur Fehlerkontrolle finden Sie bei Jensen 2012 in den Abschnitten 2.2.1 „Ursachen für Datenprobleme und Planung der Datenbereinigung“ und 2.2.2 „Einzelschritte der Datenkontrolle und Datenbereinigung“.

[Zurück zum Text](#)