Trixa, Jessica; Ebel, Thomas; Harzenetter, Karoline

Hinweise zur Aufbereitung quantitativer Daten

Version 1.3

Frankfurt am Main : DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation 2019, 11 S. - (forschungsdaten bildung informiert; 4)

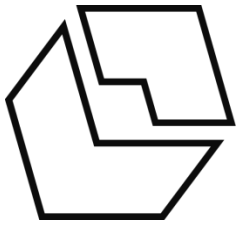
Kontakt / Contact:

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

4

forschungsdaten bildung informiert

Jessica Trixa, Thomas Ebel und Karoline Harzenetter
jessica.trixa@gesis.org // thomas.ebel@gesis.org //
karoline.harzenetter@gesis.org



Hinweise zur Aufbereitung quantitativer Daten

Version 1.3 // Februar 2019

Impressum

forschungsdaten bildung informiert // Nr. 4 (2019)

In der Reihe *forschungsdaten bildung informiert* erscheinen Beiträge zu den Themen Forschungsdaten, Forschungsdatenmanagement und Forschungsdateninfrastruktur. Publikationen in dieser Reihe sind nicht-exklusiv, das heißt, eine Veröffentlichung an anderen Orten ist möglich.

Herausgeber
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Verbund Forschungsdaten Bildung
Rostocker Str. 6
60323 Frankfurt am Main
verbund@forschungsdaten-bildung.de

Redaktion und Layout
Alexander Schuster

Empfohlene Zitation des aktuellen Heftes: Trixa, J., Thomas, E. und Harzenetter, K. (2019) Hinweise zur Aufbereitung quantitativer Daten. In: forschungsdaten bildung informiert, Nr. 4, Vers. 1.2.

www.forschungsdaten-bildung.de

Inhalt

Vorbemerkung

Um die Nachvollziehbarkeit von Forschungsdaten für Sekundärnutzer/-innen, d. h. für nicht an der Erhebung beteiligte Forscher/-innen 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 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
- 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]
- 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] Siehe auch: GESIS Empfehlungen zu Dateiformaten unter https://www.gesis.org/angebot/archivieren-und-registrieren/datenarchivierung/vorbereitung-datenuebergabe/, abgerufen am 05.12.2018

[2] Hinweise und Beispiele für Dokumentationen finden Sie unter https://www.forschungsdaten-bildung.de/dokumentieren. Zudem kann auf die Dokumentation von Skalen in der „Datenbank zur Qualität von Schule (DaQS)“ (https://daqs.fachportal-paedagogik.de/) verwiesen werden. Zur Dokumentation von Skalen siehe auch https://www.forschungsdaten-bildung.de/doku-instrumente, abgerufen am 05.12.2018.

[3] Hinweise zum Erstellen eines Methodenberichts finden sich bspw. bei Jedinger et al. (2018), Schnell (2012: 415ff.) oder bei Watteler (2010).

[4] Empfohlen wird die Angabe der Ausschöpfungsquote nach AAPOR Standard (siehe https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx, abgerufen am 05.12.2018

2 Checkliste mit den notwendigen Arbeitsschritten 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.

Dimension	Aspekte
Variablen- und Wertebenennung	☐ 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	☐ 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	☐ 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	☐ Fehlende Werte sind definiert. ☐ Fehlende Werte sind als solche durch ein Label gekennzeichnet. ☐ Existieren verschiedene Arten von fehlenden Werten, sind diese möglichst differenziert festgehalten.
Dokumentations-materialien	☐ 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.

Datenschutz und rechtliche Aspekte	☐ 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 Teilnehmer/-innen 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 Konsistenz-prüfungen	☐ Die Filterführung ist korrekt. ☐ Die Variablenausprägungen sind plausibel. ☐ Es sind keine wild codes (Werte außerhalb des zulässigen Wertebereichs) vorhanden.
Dateinamen	☐ 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	☐ Die Dateien liegen in einem empfohlenen oder zumindest akzeptablen Format vor.

3 Erläuterungen zur Aufbereitung quantitativer Daten

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). 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) (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 (Brauns et al. 2003), oder ISCED, International Standard Classification of Education der Unesco (2018).

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 Fragennummer aus dem Fragebogen oder Hinweise zu Art oder Besonderheiten einer Variable aufgenommen werden. Beispielsweise, 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).

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 Teilnehmer/-innen führen könnten. Offene Angaben sind daher sorgfältig auf datenschutzrechtliche Probleme zu untersuchen.

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

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 lautet, 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 verschiedene Arten fehlender Werte sind u. a. „Keine Angabe“, „Weiß nicht“, „Trifft nicht zu“ (Filterführung), „Split“ (Splits der Stichprobe) und „Angabe verweigert“. Weitere Beispiele und Empfehlungen des Verbunds Forschungsdaten Bildung für die Codierung fehlender Werte finden sich in *Verbund Forschungsdaten Bildung (2019): Hinweise zur Codierung fehlender Werte in der Aufbereitung quantitativer Daten. Version 1.0, fdbinfo Nr. 6* (https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdbinfo_6.pdf)

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 Forschende 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 wird 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 Nachnutzenden, die methodischen Aspekte der Studie einzuordnen und zu bewerten, bspw. ggf. die Repräsentativität der Stichprobe für die Grundgesamtheit.

[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.

Beispiele für ausführliche Datensatzdokumentationen finden sich auf der Seite: http://www.forschungsdaten-bildung.de/aufbereitung

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.

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 Teilnehmer/-innen 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 gemäß BDSG (Bundesrepublik Deutschland 2018). Besonders sensitiv sind hierbei personenbezogene Daten. Die Mindestvoraussetzung für die Übergabe von Forschungsdaten an den VerbundFDB[10] ist die formale Anonymisierung. Formales Anonymisieren umfasst das Entfernen aller direkten Identifikatoren (Namen, Anschrift, Kontaktdaten, Registernummern etc.) aus einem Datensatz durch Verschlüsselung oder Pseudonymisierung, z. B. Codierung von Befragten (Befragten-ID). Sollen Befragte zu weiteren Befragungen (beispielsweise im Rahmen eines Panels) eingeladen werden, müssen ihre Angaben getrennt von ihren Kontaktinformationen gespeichert und verarbeitet werden. Dazu werden Betroffene durch Pseudonymisierung unkenntlich gemacht, d. h. personenbezogene Daten werden durch Codes ersetzt (z. B. Klarnamen durch Befragten-ID), die personenbezogenen Daten werden dann separat und gesichert von den Forschungsdaten aufbewahrt. Kontaktdaten und Forschungsdaten sind dann nur noch über einen Schlüsselcode (z. B. Befragten-ID) verknüpfbar. Sofern die personenbezogenen Kontaktdaten nicht für weitere Forschungsvorhaben benötigt werden, müssen diese nach Beendigung des Forschungsprojekts 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 Vergröberung 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.).

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

[8] Informationen zur informierten Einwilligung bieten Metschke/Wellbrock 2002, Häder (2009: 16ff) und Jensen (2012: 14). Vorlagen für Einwilligungserklärungen in Studienteilnahmen finden Sie ebenfalls bei Metschke/Wellbrock 2002 (Anlage 1) sowie unter https://www.forschungsdaten-bildung.de/einwilligung.

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

[10] www.forschungsdaten-bildung.de/

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]

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]

vgl. https://www.forschungsdaten-bildung.de/datei-benennung

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. Forscher/-innen sollten sich um einheitliche, in der Fachdisziplin als Standard für den entsprechenden Datentyp geltende, möglichst offene oder zumindest portierbare Formate bemühen. Welche Formate für Ihre Forschungsdaten im Einzelnen empfohlen werden, haben wir für Sie unter www.forschungsdaten-bildung.de/formate zusammengefasst.

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.

[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“.

4 Quellen

Bundesrepublik Deutschland: BDSG, Bundesdatenschutzgesetz: neue Fassung vom 25.05.2018. URL: www.gesetze-im-internet.de/bdsg_2018/, abgerufen am 05.12.2018.

Brauns, H.; Scherer, Stefani; Steinmann, Susanne (2003): The CASMIN Educational Classification in International Comparative Research, in: H.P., Jürgen; Hoffmeyer-Zlotnik; Wolf, Christof (Hrsg.): Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables, Amsterdam, 196-221.

GESIS (o. J.): Vorbereitung der Datenübergabe . URL: www.gesis.org/unser-angebot/archivieren-und-registrieren/datenarchivierung/vorbereitung-datenuebergabe/, abgerufen am 05.12.2018.

Häder, Michael (2009): Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. RatSWD – Working Paper No. 90. URL: www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf, abgerufen am 05.12.2018.

Jedinger, Alexander; Watteler, Oliver, Förster, Andre (2018): Improving the quality of survey data documentation: A total survey error perspective. Data: open access 'Data in science' journal 3 (4): 45. doi: dx.doi.org/10.3390/data3040045, abgerufen am 05.12.2018.

Jensen, Uwe (2012): Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten. GESIS-Technical Reports 2012|07. URL: www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf, abgerufen am 04.12.2018.

Metschke, Rainer; Wellbrock, Rita (2002): Datenschutz in Wissenschaft und Forschung. Materialien zum Datenschutz NR. 28 (3). Aufl. Berlin, 2002. URL: https://www.uni-muenchen.de/einrichtungen/orga_lmu/beauftragte/dschutz/regelungen/ds_wiss_und_fo.pdf, abgerufen am 05.12.2018.

UNESCO Institute of Statistics (2018): International Standard Classification of Education (ISCED). URL: uis.unesco.org/en/topic/international-standard-classification-education-isced, abgerufen am 05.12.2018

Quandt, Markus; Mauer, Reiner (2012): Sozialwissenschaften. In: Neuroth, Heike; Strathmann, Stefan; Oßwald, Achim; Scheffel, Regine; Klump, Jens; Ludwig, Jens (Hrsg.): Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme. 2012, Göttingen, S. 61-81. URL: nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor_lza_forschungsdaten_bestandsaufnahme.pdf, abgerufen am 05.12.2018.

Schnell, Rainer (2012): Survey-Interview. Methoden standardisierter Befragungen. VS Verlag: Wiesbaden.

Watteler, Oliver (2010): Erstellung von Methodenberichten für die Archivierung von Forschungsdaten. GESIS. www.gesis.org/fileadmin/upload/institut/wiss_arbeitsbereiche/datenarchiv_analyse/Aufbau_Methodenbericht_v1_2010-07.pdf, abgerufen am 05.12.2018.