

Leitfaden zur Beschreibung eines Data Dictionary



Autoren: Martin Bialke, Peter Penndorf, Daniel Fredrich, Kerstin Weitmann

Überarbeitung: Henriette Rau

Version 1.3 vom 04.08.2017

mosaic-projekt@uni-greifswald.de

Inhalt

1	Motivation	3
2	Vorbedingungen	4
3	Was soll das Data Dictionary enthalten?	4
3.1	Festlegung der Variablennamen	4
3.2	Definition der Variablenausprägungen	5
3.2.1	Datentypen bestimmen.....	5
3.2.2	Wertebereiche definieren	6
3.2.3	Gültige Werte codieren	6
3.2.4	Qualitative Missings festlegen	6
3.2.5	Abhängigkeiten prüfen.....	7
3.2.6	Berechenbare Werte identifizieren.....	8
3.2.7	Pflichtelemente prüfen	8
3.2.8	Einheiten prüfen.....	8
4	Best Practices	8
4.1	Orientierung an Standards	8
4.2	Unklarheiten erkennen und kommunizieren	9
4.3	Antwortmöglichkeiten prüfen.....	9
5	Abstimmung	9
6	Vorlage und Beispiele	10
7	Literaturverzeichnis	11

HINWEIS

Sie nutzen diesen Leitfaden für Ihre Forschung? Bitte referenzieren Sie das MOSAIC Project in Ihren Arbeiten und Manuskripten:

Bialke M*, Bahls T, Havemann C, Piegsa J, Weitmann K, Wegner T, et al.
MOSAIC. A modular approach to data management in epidemiological studies.
METHODS OF INFORMATION IN MEDICINE. 2015; 54(4):364-371.
<http://dx.doi.org/10.3414/ME14-01-0133>

1 Motivation

Zur Beantwortung der wissenschaftlichen Fragestellung einer Studie oder eines Registers werden Informationen mit Hilfe von Variablen erhoben. Variablen sollten stets atomar sein und repräsentieren u.a. Antworten eines Fragebogens, Messwerte oder auch Einflussfaktoren (z.B. Untersucher, Geräte, Messmethoden, Zeitstempel).

Vor Start eines Registers oder einer Studie muss im Studienplan klar festgelegt sein, welche Variablen zur Beantwortung der wissenschaftlichen Fragestellung erforderlich sind (vgl. Gute Epidemiologische Praxis (GEP), Leitlinie 3.5 [1]). Die umfassende Beschreibung dieser Variablen sollte auch bereits mit Blick auf die spätere Datenerhebung, Qualitätssicherung und Auswertung erfolgen. Typischerweise erfolgt diese formale Beschreibung in Form eines Data Dictionary. Diese Formalisierung kann für epidemiologische Forscher ohne fundierte technische Kenntnisse eine Herausforderung darstellen.

Warum wird das Data Dictionary benötigt?

Das Data Dictionary unterstützt den Forscher dabei, zu erhebende Daten umfassend und konsistent zu dokumentieren. Ohne ergänzende Informationen, beispielsweise zu Definitionen, zu Ausprägungen, zu Gültigkeitsbereichen, zum Erhebungskontext, zu Maßeinheiten oder zur Codierung von Variablen, ist weder eine Datenauswertung noch der Austausch von Daten (z.B. mit anderen Forschern) möglich.

Ziel des vorliegenden Leitfadens ist es, Epidemiologen und Wissenschaftlern bei der Erstellung eines Data Dictionary möglichst knapp und präzise zu unterstützen. Hierzu werden zu berücksichtigende Aspekte aufgezeigt und Empfehlungen zum Vorgehen gegeben.

Warum ist Vollständigkeit und Korrektheit des Data Dictionary von Beginn an wichtig?

Die Definition des Data Dictionary muss sorgfältig durchgeführt und abgestimmt werden, denn sie ist Grundlage und Ausgangspunkt für sämtliche Folgeschritte im Verlauf der Studie bzw. des Registers. Sollten nach Studien-/Registerstart Änderungen am Studiendatensatz oder am Data Dictionary erforderlich werden, hat dies erhebliche organisatorische und zeitliche Auswirkungen, deren Aufwand regelmäßig unterschätzt wird.

Bei Änderungen muss nach ausführlicher Begründung und Dokumentation der Forschungsdatensatz angepasst und erneut mit allen beteiligten Partnern in Bezug auf Struktur, Abhängigkeiten und Ausprägungen abgestimmt werden. Erhebungsformulare sind zu überarbeiten und Datenbanken zu aktualisieren. Dies macht erneute Systemtests erforderlich. Falls sich der Prozess der Datenerhebung ändert, sind existierende Standard Operating Procedures (SOP) anzugleichen, was in Nachschulungen und einer erneuten Zertifizierung der Untersucher resultieren kann. Zudem muss bei Änderungen einer Variablen stets sorgfältig geprüft werden, inwieweit der Erhebungskontext erhalten bleibt. Beispielsweise wirken sich Anpassungen der Messmethode einer Variablen auf die Vergleichbarkeit bereits vorhandener und zukünftig erhobener Werte aus. Dies erschwert die Datenanalyse.

2 Vorbedingungen

Bevor mit der Erstellung des Data Dictionary begonnen werden kann, muss die wissenschaftliche Fragestellung der Studie/ des Registers feststehen. Ebenso sollten erforderliche Variablen und Parameter bereits grundsätzlich bekannt sein, wie Einheiten, Abhängigkeiten und mögliche Störgrößen [2]. Sofern vorhanden, wird empfohlen, auf bereits existierende nationale und internationale Datensätze zurückzugreifen [2].

„ Alle interessierenden Variablen sollen präzise definiert und möglichst standardisiert operationalisiert werden. Für die Bestimmung sind möglichst valide und reliable Mess- und Erhebungsinstrumente einzusetzen.“
(Auszug aus der Guten Epidemiologischen Praxis, Leitlinie 3.5, [1])

3 Was soll das Data Dictionary enthalten?

3.1 Festlegung der Variablennamen

Literatur und Praxis geben unterschiedliche Empfehlungen, wie Variablennamen am günstigsten zu wählen sind. Möglichst simple Bezeichner inklusive Fragennummer und aufsteigender Nummerierung stehen selbst erklärenden Variablennamen, die einen Bezug zum Fragebogen und Inhalt der Variable herstellen, gegenüber (vgl. [3], S. 27 ff). Grundsätzlich lassen sich für den Variablennamen (Kurzname) folgende allgemeine Empfehlungen zusammenfassen:

Nutzen Sie „sprechende Bezeichner“ für Variablennamen. Der inhaltliche Bezug der Variable soll ohne ergänzende Dokumentation und Kenntnis des Kontextes hergestellt werden können.

Nutzen Sie Kurz- und Langnamen. Der Kurzname muss eindeutig sein und wird für die Spaltenbezeichnung innerhalb der Datenbank genutzt. Der Langname beschreibt die Variable genauer und wird meist in Formularen als Beschreibungsfeld bzw. Label genutzt. Ergänzend können Hilfetexte sowie ausführliche Variablenbeschreibungen dargestellt werden.

Nutzen Sie Präfixe. Präfixe unterstützen allgemein die Übersichtlichkeit und die eindeutige Identifizierung einer Variablen innerhalb des Data Dictionary. Beispielsweise kann eine Variable mit Formularpräfix, auch beim Einsatz mehrerer Formulare ohne weitere Kenntnisse direkt dem Ursprungsformular zugeordnet werden.

Nutzen Sie Suffixe. Suffixe sollten eingesetzt werden, wenn mehrere sprechende Bezeichner für ähnliche Variablen als Teil von Aufzählungen gesucht sind (z.B. risikofaktor_1, risikofaktor_2, etc.). Ebenso kommen Suffixe für Verweise auf Datentypen oder Einheiten in Betracht.

Aus Sicht der Praxis muss bei der Festlegung von Variablenbezeichnern bereits die spätere Datenverarbeitung berücksichtigt werden:

Die Länge des Variablennamen muss sich an gängigen Standards orientieren. Beachten Sie bereits vorab mögliche Limitierungen des Datenbankensystems [4] und der Auswertungssoftware.¹

¹ Beispielsweise dürfen Variablennamen in SPSS erst seit Version 12.0 maximal 64 Zeichen lang sein. Vorgängerversionen mussten mit 8 Zeichen auskommen. [11]

Vermeiden Sie Groß-Klein-Schreibung. Um Probleme bei der Verarbeitung der erfassten Daten über Betriebssystemgrenzen hinweg zu vermeiden, nutzen Sie ausschließlich Kleinschreibung.

Vermeiden Sie reservierte Bezeichner. Datenbanken und viele andere technische Systeme nutzen reservierte Bezeichner, wie „begin“, „date“, „table“, „system“, „name“, etc. Werden diese als Variablennamen verwendet, können ungewollte Nebeneffekte bei der Datenverarbeitung und Systemfehler auftreten.²

Vermeiden Sie Umlaute. Vor allem deutsche Umlaute (z.B. ä, ü, ö, etc.) erzeugen immer wieder Probleme beim Speichern und Auswerten von Daten.

Vermeiden Sie Leer- und Sonderzeichen. Variablenbezeichner, die Leerzeichen, Punkte und Sonderzeichen (z.B.: „-“, „%“, „\$“, „&“, „/“, „(“, „)“) enthalten, erschweren bereits das Anlegen der Datenbanken, da sie je nach Datenbanksystem unzulässig sind [4] und Systemfehler verursachen können. Die Verbindung von Wörtern mittels Unterstrich „_“ ist dagegen möglich.

Achten Sie auf internationale Anschlussfähigkeit. Nutzen Sie nach Möglichkeit englischsprachige Variablenbezeichner, um Verständnis und Aussagekraft der Variablen für internationale Auswertungen und Veröffentlichungen zu unterstützen.

Geben Sie Strukturinformationen. Bezeichner von Fragebogenvariablen sollten idealerweise Verweise auf die Fragebogenstruktur (Kapitel- oder Abschnittsnummern) enthalten. Diese Strukturinformation wird benötigt, wenn nur Teile der erhobenen Datensätze für die Datenauswertung exportiert werden.

Tabelle 1 Beispiele für "schlechte" Variablenbezeichner

Variablenname	Problem	Verbesserter Variablenname (Kurzname)
Zw-dauer (h)	Groß-Klein-Schreibung, Sonderzeichen, Leerzeichen	frm_aufnahme_zuweisungsdauer_stunden
n1, n2, n3	Kein Hinweis auf Variableninhalt	frm_aufnahme_unfallursache1, frm_aufnahme_unfallursache2, frm_aufnahme_unfallursache3
datum1	nicht „sprechend genug“	frm_aufnahme_datum

3.2 Definition der Variablenausprägungen

3.2.1 Datentypen bestimmen

Fast jede zu erhebende Variable lässt sich bei Bedarf durch elementare Datentypen abbilden:

Ganzzahlen sind neben der Angabe von Häufigkeiten vor allem für die Abbildung kategorialer Daten geeignet, z.B. codierte Auswahlmöglichkeiten zur Ursache eines Unfalls.

Gleitkommazahlen sind vor allem für metrische Werte geeignet, z.B. die Angabe der Blutalkoholkonzentration. Hier sollten stets die erforderlichen Nachkommastellen festgelegt werden.

Zeichenketten eignen sich für Texte. Die maximale Länge der Zeichenkette sollte stets vorgegeben werden, z.B. „Beschreibung des Unfallhergangs (max. 255 Zeichen)“.

² Eine Liste reservierter Bezeichner am Beispiel von MySQL finden Sie unter folgender Adresse: <http://dev.mysql.com/doc/refman/5.6/en/keywords.html>

Datum und Uhrzeit lassen sich je nach Erhebungswerkzeug und Datenbanksystem unterschiedlich abbilden. Grundsätzlich ist eine Abbildung in Form von Zeichenketten möglich. In jedem Fall sollte für Datums- und Zeitangaben das zu nutzende Format einheitlich definiert werden, z.B. „hh:mm“ oder „dd.mm.yyyy“.

3.2.2 Wertebereiche definieren

Für jede Variable sollte ein gültiger Wertebereich definiert werden, um bereits bei der Datenerhebung die Eingabe nicht plausibler Werte zu minimieren. Der Wertebereich muss alle real vorkommenden Werte umfassen und darf nicht zu knapp bemessen sein.

Geben Sie nach Möglichkeit für die spätere Qualitätssicherung und Auswertung das Skalenniveau³ an. Dies erlaubt die Unterscheidung von Wertebereichen und unterstützt so die Datenauswertung. Beispielsweise wird die Variable "Unfallursache" (Wertebereich 1..5 ganzzahlig, nominal) in anderer Weise ausgewertet als z.B. eine Variable "Körpergröße", die zwar ebenfalls ganzzahlig angegeben werden kann, jedoch metrisch ist und somit u.a. Wertevergleiche und Mittelwerte erlaubt.

Tabelle 2 Beispiele für Zielgrößen und mögliche Wertebereiche

Zielgröße	Möglicher Wertebereich	Skalenniveau
Auswahlmöglichkeiten 1, 2 und 3	1,2,3	nominal
Angabe in Prozent	0 - 100	kardinal
Größe in cm	1 - 299	kardinal
Aufnahmedatum	01.01.2015 - 31.12.2015	kardinal

3.2.3 Gültige Werte codieren

Werden Fragen durch die Auswahl einer oder mehrerer vorgegebener Antworten bzw. Antwortkategorien beantwortet, ist jede der möglichen Antworten eindeutig zu codieren. Diese Codierungen müssen eindeutig, einheitlich und erweiterbar sein, die vollständige Antwortmenge abbilden und sich gegenseitig ausschließen. Dazu werden numerische bzw. alphanumerische Codes definiert [3].

Tabelle 3 Beispiele für Codierung gültiger Werte. Die Zuordnung von Wert und Codierung erfolgt hier basierend auf der Reihenfolge der Werte und Codierungen.

Beispielvariable	Mögliche Werte	Mögliche Codierung
Standardfrage	Nein, Ja	0,1
Geschlecht	Männlich, weiblich	m, f
Unfallursache	Verbrühung, Flamme, Explosion, Säure, Lauge	1,2,3,4,5

Weitere Beispiele und Tipps sind u.a. in den *Leitlinien zum Forschungsdatenmanagement* [3] (S.29 ff) zu finden.

3.2.4 Qualitative Missings festlegen

Grundsätzlich sollten stets auch fehlende Informationen (Missings) so genau wie möglich beschrieben werden. Wird dies vernachlässigt, kann in der späteren Auswertung der erhobenen

³ Beispiele für Skalenniveaus unter <http://www.reiter1.com/Glossar/Skalenniveaus.htm>, Zugriff am 15. September 2015

Datensätze keine Aussage gemacht werden, ob eine Variable fälschlicherweise nicht erfasst wurde oder das Fehlen bedingt ist durch z.B.:

- fehlende Information (Befragter weiß es nicht)
- fehlende Bereitschaft (Befragter verweigert Antwort)
- Nicht-Abdeckung durch Fehler im Erhebungsinstrument (Fragestellung /Antwortmöglichkeit nicht zutreffend)
- Frage nicht gestellt / Untersuchung nicht durchgeführt
- fehlende Dokumentation (Ergebnis wurde nicht notiert)

Um in der Auswertung qualitative Aussagen über fehlende Werte machen zu können und somit die Vollständigkeit der Datengrundlage der statistischen Analysen und Berichte belegen zu können [2], müssen alle Missings codiert werden.

Achten Sie auf Einheitlichkeit und berücksichtigen Sie bei der Codierung der qualitativen Missings die Wertebereiche gültiger Werte. Die *Leitlinien zum Forschungsdatenmanagement* bieten dazu zahlreiche Tipps und praktische Beispiele (vgl. [3] S. 31 ff):

Nutzen Sie Missing-Codes außerhalb des gültigen Wertebereiches: Setzen Sie die höchsten numerischen Codes ein, die sich außerhalb des gültigen Wertebereichs sinnvoll darstellen lassen (z.B. 99977=nicht erhoben, 99978=keine Angabe, etc.) oder verwenden Sie negative Missing-Codes (-1, -2, etc.).

Alternativ kann die Codierung der qualitativen Missings u.a. auf Basis des *HL7 Null Flavour Standards* [5] oder der *Codierung von Null Values in OpenClinica* [6] erfolgen.

3.2.5 Abhängigkeiten prüfen

Darf eine Variable nur erfasst werden, wenn eine andere Variable einen bestimmten Wert annimmt, so muss diese Abhängigkeit klar benannt und die daraus resultierende Bedingung eindeutig definiert werden.

Tabelle 4 Beispiele für Abhängigkeiten von Variablen

<i>Variable A</i>	<i>Variable B</i>	<i>Abhängigkeit</i>
frm1_accident_cause_11 (Unfallursache: Sonstiges)	frm1_accident_description (Beschreibung des Unfall- zusammenhangs)	Die Beschreibung des Unfallzusammenhangs darf nur dann erfolgen, wenn als Unfallzusammenhang „Sonstiges“ ausgewählt wurde.
frm1_krea_mgdl (Kreatin-Wert in mg/dl)	frm1_krea_umoll (Kreatin-Wert in μ mol/l)	Die gleichzeitige Angabe des Kreatin-Wertes in beiden Einheiten (mg/dl und μ mol/l) ist unzulässig.
frm1_admission_date (Aufnahmedatum)	frm1_release_date (Entlassungsdatum)	Das Entlassungsdatum muss dem Datum der Aufnahme oder einem späteren Datum entsprechen.

Typische Beispiele sind geschlechtsspezifische Fragen: Männer müssen Fragen zur Schwangerschaft nicht beantworten und können entsprechende Fragen im Fragebogen „überspringen“. Diese Sprünge werden durch Sprungvariablen im Data Dictionary kenntlich gemacht. Durch die Verwendung von

Sprungvariablen können die Anforderungen der späteren Auswertung, der Qualitätssicherung und der Plausibilitätsprüfung bereits während der Datenerhebung berücksichtigt und unterstützt werden.

3.2.6 Berechenbare Werte identifizieren

Automatische Berechnungen von Variablen, wie beispielsweise des Body Mass Index (BMI) aus Größe und Gewicht, können den Erhebungsaufwand reduzieren und das Auftreten von Inkonsistenzen durch Berechnungsfehler vermindern. Andererseits können fehlerhafte Formeln die Qualität der Daten systematisch beeinträchtigen. Die Notwendigkeit zur Prüfung und Validierung der Eingaben bleibt also bestehen.

Identifizieren Sie berechenbare Variablen: Gehen Sie von den Variablenabhängigkeiten aus und spezifizieren Sie die erforderlichen Formeln.

3.2.7 Pflichtelemente prüfen

Grundsätzlich sollte der Studiendatensatz nur die für die Beantwortung der wissenschaftlichen Fragestellung der Studie/ des Registers erforderlichen Variablen enthalten. Diese Pflichtelemente stellen das **Minimal Dataset** (Kern-Datenelemente [2]) dar.

Werden darüber hinaus weitere Variablen erfasst, sollten Pflichtangaben und optionale Angaben eindeutig im Studiendatensatz gekennzeichnet werden. Die Verwendung von optionalen Variablen birgt jedoch die Gefahr des unvollständigen Ausfüllens und kann bei der späteren Auswertung zu Problemen führen [2].

3.2.8 Einheiten prüfen

Um das Fehlerpotential bei der Datenauswertung zu reduzieren, müssen sämtliche Einheiten je Variable eindeutig spezifiziert werden. Dabei sind folgende Aspekte zu berücksichtigen:

Nutzen Sie typische Einheiten: Messwerte werden je Anwendungskontext durchaus in unterschiedlichen Einheiten erfasst. Orientieren Sie sich bei der Wahl der Einheiten an der Praxis und gängigen Standards, z.B. den internationalen SI-Einheiten (in der jeweils aktuellen Fassung) [7].

Achten Sie auf Einheitlichkeit der Angaben: Nach Möglichkeit sollten ähnliche Zielgrößen in gleichen Einheiten angegeben werden, z.B. Größen- und Längenangaben stets in „cm“. Vermeiden Sie variierende Angaben, wie „%“ und „Prozent“.

4 Best Practices

4.1 Orientierung an Standards

Gemäß der GEP [1] sollte bei der Beschreibung der Variablen und deren Ausprägungen nach Möglichkeit auf gängige nationale und internationale Standards zurückgegriffen werden. Auf diese Weise kann die Qualität und Vergleichbarkeit der Daten aus epidemiologischen Studien und Registern unterstützt werden.

Im Kontext epidemiologischer Studien und Register können etablierte Standards zur Klassifizierung von Diagnosen und Krankheiten (z.B. ICD [8], DIMDI), zur Beschreibung von Variablen und Ausprägungen (z.B. LOINC [9]) sowie zur Beschreibung von Metadaten (z.B. Nationales Metadaten Repository [10]) die erforderliche Orientierung bieten.

4.2 Unklarheiten erkennen und kommunizieren

Die Definition von Variablen erfordert auch bei guter Vorbereitung und Planung ein erhebliches Maß an Kommunikation aller Beteiligten untereinander in Bezug auf Erforderlichkeit und Angemessenheit von Zielgrößen, Wertebereichen, Datentypen und Abhängigkeiten.

Um den bereits in Abschnitt 1 skizzierten Mehraufwand durch nachträgliche Anpassungen auf ein Mindestmaß zu reduzieren, sollten auftretende Unklarheiten bei der Festlegung des Data Dictionary zum frühestmöglichen Zeitpunkt dokumentiert, kommuniziert und in Zusammenarbeit mit allen Beteiligten geklärt und beseitigt werden.

4.3 Antwortmöglichkeiten prüfen

Vermeiden Sie die Möglichkeit von Mehrfachantworten je Fragestellung, beispielsweise um mehrere Ursachen eines Unfalls aufzuzählen. Diese meist konkateniert (z.B. komma-separiert) abgespeicherten Werte erschweren die Datenauswertung und reduzieren die Qualität der gespeicherten Daten. Überdenken Sie im Einzelfall die entsprechende Fragestellung.

Um Daten effektiv analysieren zu können, sollte eine Variable stets nur einen einzigen Wert enthalten. Führen Sie alternativ je Antwortoption eine weitere Variable ein (z.B. frm1_ursache1, frm1_ursache2, etc.) und erfassen Sie, ob diese „zutreffend“ oder „nicht zutreffend“ ist.

Vermeiden Sie zudem nach Möglichkeit die Angabe von nicht-kategorisierten Freitexten. Diese erhöhen erheblich den Aufwand für Qualitätssicherung und Auswertungen, da sie praktisch nicht automatisiert auswertbar sind.

5 Abstimmung

Die mit Hilfe dieses Leitfadens entstehende Beschreibung des Data Dictionary soll Grundlage für Formularentwürfe⁴ und ein Datenmodell sein.

Um unnötigen Arbeitsaufwand zu vermeiden, sollte nach Erstellung des Data Dictionary eine Rücksprache mit möglichst allen beteiligten Forschern unter Einbeziehung von Methodikern zur Qualitätskontrolle erfolgen. Zu prüfen ist, ob der Zieldatensatz eine ausreichende Analysetiefe und Qualität besitzt [2] und ob damit eine wissenschaftlich korrekte Beantwortung der Studien-/Registerfragestellung gewährleistet werden kann.

Rat und Hilfestellung bei der Planung, Durchführung und Auswertung von epidemiologischen Studien und Registern bieten die Arbeitsgruppen der *Deutschen Gesellschaft für Epidemiologie e.V. (DGEpi)*, des *Deutschen Netzwerks für Versorgungsforschung e.V. (DNVF)* und der *Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF)*.

⁴ Beachten Sie bitte dazu den MOSAIC Leitfaden zur Erstellung von eCRFs.

6 Vorlage und Beispiele

Zur formalen Beschreibung des Data Dictionary wird eine [Vorlage](#) im Excel-Format als Download bereitgestellt.

https://mosaic-greifswald.de/fileadmin/Produkte/Leitfaden_DataDictionary/Vorlage2_v2_ger.xlsx

Nutzen Sie die Vorlage gern als Arbeitsgrundlage und erweitern Sie diese nach Bedarf. Anmerkungen, Ergänzungen und Fragen dazu bitte per Mail an mosaic-projekt@uni-greifswald.de.

7 Literaturverzeichnis

- [1] German Society of Epidemiology (DGEpi), **Guidelines and Recommendations to Assure Good Epidemiologic Practice (GEP)**, 2008. [Online]. URL: http://dgepi.de/fileadmin/pdf/GEP_LL_english_f.pdf. [Zugriff am 24 2 2014].
- [2] D. Müller, M. Augustin, N. Banik, W. Baumann, K. Bestehorn, J. Kieschke und R. Lefering, „**Memorandum Register für die Versorgungsforschung**,“ *Gesundheitswesen 2010*, Bd. 72, S. 824-839, 2010.
- [3] U. Jensen, **Guidelines for the management of research data (Leitlinien zum Management von Forschungsdaten)**, GESIS - Leibniz-Institut für Sozialwissenschaften, Köln, 2012.
- [4] **MySQL-Homepage (Schema Object Names)**, 2015. [Online]. URL: <https://dev.mysql.com/doc/refman/5.0/en/identifiers.html>. [Zugriff am 16 7 2015].
- [5] FHIR Project, **Value Set for Codes in HL7 Null-flavors**, 2014. [Online]. URL: <http://hl7.org/implement/standards/fhir/null-flavor.html>. [Zugriff am 24 2 2015].
- [6] Open Clinica, **Null-Value Codierung in Open Clinica**, [Online]. URL: <https://community.openclinica.com/OpenClinica/3.0/doc/glossary.html#nullValues>. [Zugriff am 24 2 2015].
- [7] Bureau International des Poids et Mesures, **SI Brochure: The International System of Units (SI), 8th Edition**, 2014. [Online]. URL: http://www.bipm.org/utils/common/pdf/si_brochure_8.pdf. [Zugriff am 24 8 2015].
- [8] World Health Organization, **International Statistical Classification of Diseases and Related Health Problems, 10th Revision**, 2010. [Online]. URL: <http://apps.who.int/classifications/icd10/browse/2010/en>. [Zugriff am 12 4 2015].
- [9] Regenstrief Institute, Inc., **LOINC.org**, 2015. [Online]. URL: <https://loinc.org/>. [Zugriff am 12 4 2015].
- [10] Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE), **Nationales Metadata Repository**, 2015. [Online]. URL: <https://mdr.imise.uni-leipzig.de/>. [Zugriff am 12 4 2015].
- [11] S. Boslaugh, **An Intermediate Guide to SPSS Programming: Using Syntax for Data Management**, 1. Hrsg., SAGE Publications, 2004.