

Juli 2010

Erstellung von Methodenberichten für die Archivierung von Forschungsdaten

Oliver Watteler

1. Erstellung von Methodenberichten für die Archivierung von Forschungsdaten

Ein Methodenbericht stellt, neben Erhebungsinstrumenten und Datensätzen, die dritte, zentrale Einheit bei der Archivierung empirischer Studien dar. Ein Methodenbericht sollte dem Forscher einen Überblick über die Erhebung und die Aufbereitung der zu analysierenden Daten bieten.

Leitfragen bei der Erstellung eines solchen Berichts sind dabei:

- | | |
|--|---|
| a) Was wurde erhoben oder erfasst? | Art der Daten, technisches Format |
| b) Wozu wurden die Daten erhoben? | Zweck der ursprünglichen Forschung |
| c) Wer war an der Datenerhebung beteiligt? | Primärforscher, Feldforschungseinrichtung etc. |
| d) Wann wurden die Daten erhoben? | Feldphase |
| e) Wie wurden die Daten erhoben? | Auswahl der Untersuchungseinheiten, Modus der Erhebung, Aufbau des Erhebungsinstrumentes oder -schemas etc. |
| f) Wo fand die Datenerhebung statt? | Regionale Ausdehnung, Ort der Befragung oder Erhebung |

Die Leitfragen sollen einen Eindruck der Daten und ihrer Qualität vermitteln. Der Begriff der Qualität wird dabei teilweise sehr weit gefasst und soll hier nicht weiter definiert werden. Generell kann jedoch festgehalten werden, dass eine ausreichende Dokumentation dem Forscher bei einer Sekundäranalyse die Möglichkeit bieten sollte, die Datenerstellung und damit die Basis für veröffentlichte Analysen nachvollziehen zu können (intersubjektive Überprüfbarkeit). Bei der Datenerhebung – unser Fokus liegt auf Umfragedaten – kann es z.B. zu Ausfällen beim Sampling, Überrepräsentation von Teilpopulationen oder sog. Modeeffekte kommen, die bei weiteren Auswertungen eventuell korrigiert werden können.

Biemer/Lyberg 2003 sowie Weisberg 2005 bieten Überblicke über die Faktoren, die die Datenerhebung bei Umfragen beeinflussen können. Bereits 1999 veröffentlichte die Deutsche Forschungsgemeinschaft einen Katalog von Qualitätskriterien für die Umfrageforschung (DFG 1999) und 2006 wurde ein entsprechender internationaler Standard für die Markt- und Meinungsforschung verabschiedet.

Die folgenden Hinweise zur Erstellung eines Methodenberichtes orientieren sich in erster Linie am Bericht der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ ALLBUS (Wasmer e.a. 2010) und werden durch Beispiele aus anderen Beschreibungen ergänzt. Sie sollen Forschern und Projekten Anregungen und Hilfestellungen bieten.

2. Checkliste für Methodenbericht

Die Checkliste für einen Methodenbericht kann parallel zum Forschungsprozess in vier Abschnitte unterteilt werden: Konzept, Sampling, Datenerhebung und Datenaufbereitung. Diese vier Abschnitte werden in Methodenberichten unterschiedlich benannt und nicht immer in dieser Reihenfolge aufgeführt. Auch variiert der Umfang der Dokumentation sehr stark, da z.B. nicht immer alle Details des Samplingprozesses vom Feldinstitut an die Forscher gemeldet werden. D.h. die Liste bildet einen Idealfall ab. Parallel können viele der im Methodenbericht zusammengestellten Informationen in eine Studienbeschreibung eingehen, die Teil von Katalogen wie dem [Datenbestandskatalog](#) des Archivs von GESIS sind.

1) Konzept (auch: Studiendesign, Anlage)

	Bezeichnung	Alternative Bezeichnung	Inhalt
1.1	Forschungsziel	Überblick	Übergeordnetes Ziel der Studie: Was sollte untersucht werden?
1.2	Studiendesign		Anlage der Studie: Wie sollten Forschungsfragen beantwortet oder Thesen überprüft werden?
1.3	Befragungsmethode	Erhebungsmethode	Wie sollten die Daten erhoben
1.4	Frageprogramm		Liste der Fragenelemente, die an die Untersuchungsperson gerichtet wurden. Enthalten im Erhebungsinstrument (= Fragebogen).

2) Sampling (auch: Stichprobe, Auswahl)

	Bezeichnung	Alternative Bezeichnung	Inhalt
2.1	Sampling	Stichprobeanlage	Spezifikation, nach welchem Verfahren die Stichprobe gezogen wurden (z.B. mehrfachgeschichtete Zufallsstichprobe nach ADM-Design)
2.2	Grundgesamtheit	Sampling Frame, Universum	Auf welche Gruppe von Untersuchungseinheiten bezieht sich die Untersuchung. Sampling Frame stellt, aufgrund bekannter Probleme von Nonresponse bei „repräsentativen“ Erhebungen, die Alternative zu einer vollständigen Grundgesamtheit dar.
2.3	Auswahl	Stichprobenrealisierung	Wie wurde die Auswahl genau durchgeführt?
2.4	Stichprobenkorrekturen		Welche Korrekturen waren eventuell notwendig (z.B. Nachziehung von Adressen oder Telefonnummern)?
2.5	Gewichtung		Welche Gewichtungsvariablen wurden auf der Basis des Auswahlverfahrens wie berechnet und wozu dienen sie?

3) Feldphase (auch: Datenerhebung, Befragung)

3.1) Pretest

Zweck eines Pretests ist i.d.R. die eigentliche Erhebung kritisch vorzubereiten, Erhebungsinstrumente zu testen, Auswahlverfahren zu verfeinern etc. D.h. es können hier wichtige Weichenstellungen erfolgen, die von methodischem Interesse sind, wie etwa die Änderungen von Frageformulierungen, der Austausch von Interviewern (sensible Themen bei denen z.B. das Geschlecht der Interviewer einen Einfluss haben kann), die Kürzung des Erhebungsinstruments etc.

	Bezeichnung	Alternative Bezeichnung	Inhalt
3.1.1	Pretest		Mode der Datenerhebung (CATI, CAPI, Paper&Pencil, etc.). Wurde der Pretest unter den selben Bedingungen durchgeführt wie die eigentliche Erhebung?
3.1.2	Interviews	Kontakte	Welche Besonderheiten gab es bei der Datenerhebung? (z.B. Anwesenheit Angehöriger bei Face-to-face Befragungen)
3.1.3	Interviewer		Welche Interviewer mit welcher Qualifikation wurden eingesetzt?
3.1.4	Ergebnisse		Änderungen im Studiendesign als Folge des Pretests.

3.2) Eigentliche Erhebung

	Bezeichnung	Alternative Bezeichnung	Inhalt
3.1	Befragung	Erhebung	Mode der Datenerhebung (CATI, CAPI, Paper&Pencil, etc.)
3.2	Interviews	Kontakte	Welche Besonderheiten gab es bei der Datenerhebung? (z.B. Anwesenheit Angehöriger bei Face-to-face Befragungen)
3.3	Interviewer		Welche Interviewer mit welcher Qualifikation wurden eingesetzt?
3.4	Ausschöpfung	Rücklauf	Wie wurden die Untersuchungspersonen rekrutiert, d.h. wie groß war die ursprünglich geplante Auswahlmenge (Bruttostichprobe), welche Ausfälle gab es und wie groß war die Gruppe der Untersuchungspersonen (Nettostichprobe). Der Quotient aus Brutto- und Nettostichprobe bildet die Ausschöpfungsquote. (Vgl. Anhang 1)

4) Datenaufbereitung

	Bezeichnung	Alternative Bezeichnung	Inhalt
4.1	Datenerfassung		Wie wurde der vorliegende Datensatz erstellt? Welche Korrekturen wurden vorgenommen? Welchen Vorgaben und Standards wurde gefolgt (z.B. Berufsvercodung nach ISCO ¹ oder Bildungsvercodung nach ISCED ²)?
4.2	Datenprüfung	Qualitätskontrolle	Kontrolle der Reliabilität, Validität und Repräsentativität. Maßnahmen der Qualitätskontrolle können sein z.B. die Schätzung fehlender Angaben, der Vergleich mit Daten des Mikrozensus (Verlässlichkeit der Stichprobe), die Berechnung von Konfidenzintervallen etc.
4.3	Anonymisierung	Datenschutz	Maßnahmen zum Schutz der Befragten im Datensatz. Einige Samples (z.B. Eliten) machen es notwendig, Daten, die an Dritte für Sekundäranalysen weitergeleitet werden, in ihrem Informationsgehalt zu verändern, um eine Re- oder Deanononymisierung der Befragungseinheiten zu verhindern.

¹) ISCO = International Standard Classification of Occupations

²) ISCED = International Standard Classification of Education

Anhang 1: Tabelle für die Darstellung der Ausschöpfung

		Befragte	
		N	%
	Ursprüngliche Bruttostichprobe		
+	Ersatz für stichprobenneutrale Ausfälle		
=	Bruttostichprobe		
-	Stichprobenneutrale (Qualitätsneutrale) Ausfälle		
=	Bereinigte Stichprobenansatz		
-	Systematische Ausfälle		
=	Auswertbare Interviews (Nettostichprobe)		

(Quelle: Wasmer e.a. 2010: 56)

Anhang 2: Schema der Studienbeschreibung Datenarchiv GESIS

(Beispiel: ALLBUS 2008)

Feldname	Erläuterung
ZA-Studiennummer 4600	[Internes Feld; Kann z.B. durch das Kürzel für eine Welle ersetzt werden.]
Titel Allgemeine Bevölkerungsumfrage der Sozialwissenschaften ALLBUS 2008	Vollständiger Titel der Studie [Kann von Projekttiteln abweichen!]
Erhebungszeitraum März 2008 bis August 2008	
Primärforscher/Wissenschaftlicher Beirat, Institution Hans-Jürgen Andreß Universität Köln, Vorsitz; Andreas Diekmann ETH Zürich; Hubert Feger Freie Universität Berlin; Stefan Liebig Universität Duisburg, Bielefeld; Heiner Meulemann Universität Köln; Rüdiger Schmitt-Beck Universität Duisburg, Mannheim; Heike Trappe Universität Rostock	Alle für die Vorbereitung und Konzeptionierung verantwortlichen Forscher und Einrichtungen.
Datenerhebung TNS Infratest Sozialforschung, München	Die für die eigentliche Datenerhebung verantwortliche Stelle.
Inhalt 1. Wirtschaft 2. Mediennutzung 3. Politische Einstellungen und politische Partizipation 4. Soziales Kapital 5. Politik und soziale Ungleichheit 6. Staatsbürgerschaften und Herkunftsland 7. Nationalstolz 8. Sonstiges 9. ALLBUS-Demographie 10. Daten zum Interview 11. Freizeit und Sport (ISSP) 12. Religion III (ISSP) 13. Abgeleitete Indizes	Kurze Darlegung des Inhaltes der Studie. Im Falle des Datenarchivs von GESIS eine Beschreibung aller verwendeter Items sowie aller im Datensatz enthaltenen Variablen.

Grundgesamtheit Untersuchungsgebiet: Bundesrepublik Deutschland	[Universe, Sampling Frame]
Auswahl Personenstichprobe: Zweistufige, disproportional geschichtete Zufallsauswahl in West- (incl. West-Berlin) und Ostdeutschland (incl. Ost-Berlin) aus allen deutschsprachigen Personen, die zum Befragungszeitpunkt in Privathaushalten lebten und vor dem 01.01.1990 geboren sind. In der ersten Auswahlstufe wurden Gemeinden in Westdeutschland und in Ostdeutschland mit einer Wahrscheinlichkeit proportional zur Zahl ihrer erwachsenen Einwohner ausgewählt. In der zweiten Auswahlstufe wurden Personen aus den Einwohnermeldekarteien zufällig gezogen.	Beschreibung des Auswahlverfahrens der Untersuchungseinheiten.
Erhebungsverfahren Mündliche Befragung mit standardisiertem Frageprogramm (CAPI – Computer Assisted Personal Interviewing), eine Fragebatterie zur politischen Informiertheit als CASI (Computer Assisted Self-Interviewing).	
Datensatz Daten-Typ SPSS Anzahl der Einheiten 3469 Anzahl der Variablen 800 Analyse-System(e) SPSS, Stata	Format und Umfang des Datensatzes.
Veröffentlichungen	
Weitere Hinweise Befragte aus dem Bereich der neuen Bundesländer sind in den Daten überrepräsentiert (oversample).	

Referenzen

Biemer/Lyberg 2003: Paul B. Biemer, Lars Lyberg, Introduction to survey quality, Hoboken

DFG 1999: Deutsche Forschungsgemeinschaft, Qualitätskriterien der Umfrageforschung. Denkschrift, Herausgeg. von Max Kaase, Berlin

ISO 20252-2006: ISO 20252-2006, Markt-, Meinungs- und Sozialforschung - Begriffe und Dienstleistungsanforderungen

Wasmer e.a. 2010: Martina Wasmer, Evi Scholz, Michael Blohm, Konzeption und Durchführung der ‚Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften‘ (ALLBUS) 2008, Mannheim (= [GESIS-Technical Reports 2010|04](#)) (15.07.2010)

Weisberg 2005: Herbert F. Weisberg, The Total Survey Error Approach. A Guide to the New Science of Survey Research, Chicago u.a.