

Elke Middendorff / Marten Wallis

17. – 21. Sozialerhebung 2003 – 2016

Daten- und Methodenbericht zum gepoolten Datensatz
der fünf Studierendenbefragungen

Daten- und Methodenbericht

Juni 2021

Autor*innen:
Dr. Elke Middendorff
Martens Wallis

Der vorliegende Bericht soll folgendermaßen zitiert werden:
Middendorff, E. & Wallis, M. (2021). 17. – 21. Sozialerhebung. Daten- und Methodenbericht zum gepoolten Datensatz der fünf Studierendenbefragungen 2003 – 2016. Version 1.0.0. Hannover: FDZ-DZHW.

Herausgeber:
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)
Lange Laube 12 | 30159 Hannover | www.dzhw.eu
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960 | info@dzhw.eu

Geschäftsführung:
Prof. Dr. Monika Jungbauer-Gans
Karen Schlüter

Vorsitzender des Aufsichtsrats:
Ministerialdirigent Peter Greisler

Registergericht:
Amtsgericht Hannover | B 210251

Dieses Werk steht unter der Creative Commons "Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz" (CC-BY-NC-SA)
<https://creativecommons.org/licenses/by-nc-sa/3.0/>



Inhaltsverzeichnis

Tabellenverzeichnis	2
1 Einleitung	3
2 Datennutzungshinweise	4
3 Harmonisierung.....	5
4 Schema der Variablennamen	8
4.1 Variablennamenschemata der einzubeziehenden Datensätze	8
4.2 Variablennamenschema des gepoolten Datensatzes	9
4.3 Das Suffix.....	11
4.4 Systematik fehlender Werte	12
5 Variablen im Datensatz	14
5.1 Sortierung der Variablen im Datensatz.....	14
5.2 Übersicht zur sozialerhebungsbezogenen Variablen-Präsenz	14
Literaturverzeichnis	16

Tabellenverzeichnis

Tabelle 1: Standards und Zugangswege der SUFs zur 17. – 21. Sozialerhebung.....	8
Tabelle 2: Variablennamenschemata der SUFs zur 17. – 21. Sozialerhebung.....	9
Tabelle 3: Teilelemente und Zusammensetzung des Variablenstammes	10
Tabelle 4: Themengebiete in den Variablennamen des gepoolten Datensatzes (17. – 21. Sozialerhebung 2003 – 2016)	10
Tabelle 5: Zusammenfassende Übersicht der verwendeten Suffix Kürzel	11
Tabelle 6: Systematik fehlender Werte in den Quell-Datensätzen (17. – 21. Sozialerhebung)	12
Tabelle 7: Systematik fehlender Werte im gepoolten Datensatz der 17. – 21. Sozialerhebung.....	13
Tabelle 8: Übersicht zur sozialerhebungsbezogenen Variablen-Präsenz.....	15

1 Einleitung

Die Sozialerhebung des Deutschen Studentenwerks (DSW) ist eine seit 1951 bestehende Untersuchungsreihe zur wirtschaftlichen und sozialen Lage der Studierenden in Deutschland.¹ Sie wird seit 1982² (10. Sozialerhebung) im Auftrag des bzw. seit der 21. Sozialerhebung in Kooperation mit dem DSW durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)³ durchgeführt. Das Bundesministerium für Bildung und Forschung (BMBF) fördert die Studie seit der 6. Sozialerhebung (1967/1968). Die Sozialerhebung dient – in Ergänzung zur amtlichen Hochschulstatistik – unter anderem dem nationalen und internationalen Bildungsmonitoring. Darüber hinaus liefert sie wichtiges Steuerungswissen für hochschul- und sozialpolitische Fragen sowie belastbare und umfassende Daten für die Forschung.

Im Rahmen der Tätigkeit des vom BMBF geförderten Forschungsdatenzentrums für Hochschul- und Wissenschaftsforschung am DZHW (FDZ-DZHW) werden die Daten einiger jüngerer Erhebungen dieser Reihe nachträglich zum Zweck der Nachnutzung aufbereitet und dokumentiert.⁴ Bisher stehen die 17., 18., 19., 20. und 21. Sozialerhebung als einzelne *Scientific Use Files* (SUF) für die wissenschaftliche Sekundärnutzung zur Verfügung. Neben dem Datensatz der einzelnen Erhebung wird jeweils auch Dokumentationsmaterial zum Datensatz und zur Durchführung der Studien bereitgestellt.⁵ Diese fünf Einzeldatensätze der 17. – 21. Sozialerhebung wurden in einen gemeinsamen Datensatz gepoolt und stehen jetzt als Scientific Use File (ssypool-SUF) sowie als Campus Use File (ssypool-CUF) zum Download bereit.

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zum gepoolten Datensatz (doi: [10.21249/DZHW:ssypool:1.0.1](https://doi.org/10.21249/DZHW:ssypool:1.0.1)).⁶ Weitere Dokumentationsmaterialien zum Datenpaket sind der Datensatzreport, die Fragebogen und Variablen-Fragebogen der fünf einbezogenen Sozialerhebungen sowie die Dokumentation zur Variablen-Harmonisierung inklusive einer Übersicht der enthaltenen Variablen. Alle Materialien sind zur Ansicht und zum Download im Metadaten-System des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) frei zugänglich.

Die zentralen Informationen zur Nutzung des Datensatzes folgen im 2. Kapitel. In Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.** werden die bei der Harmonisierung

¹ Weiterführende Informationen, alle Berichte und Materialien zur Sozialerhebung stehen auf der Website des Projekts zur Verfügung (<http://www.sozialerhebung.de>).

² Die 1. (1951) und 2. Sozialerhebung (1953) wurden vom Studentenwerk Frankfurt am Main im Auftrag des Verbands Deutscher Studentenwerke durchgeführt. Das Studentenwerk Frankfurt am Main führte auch die 3. (1956) bis 9. Sozialerhebung (1979) durch, die vom Deutschen Studentenwerk (DSW) beauftragt wurden. Einen detaillierten Überblick über Akteure, Methoden, Themen und projektbezogene Publikationen der Untersuchungsreihe von ihren Anfängen bis zur 21. Sozialerhebung bietet ein Working Paper von Middendorff 2019 (<http://www.sozialerhebung.de/Hintergrund/geschichte>).

³ Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung (DZHW, <http://www.dzhw.eu>) entstand im August 2013 durch eine Ausgründung aus der HIS Hochschul-Informations-System GmbH. Im nachfolgenden Text wird stets der Begriff DZHW verwendet, auch wenn die Studie vor der Ausgründung durchgeführt wurde. Auf allen Originaldokumenten der 17. bis 19. Sozialerhebung (Fragebogen, Flyer etc.) sowie in den dazugehörigen Berichten ist entsprechend die HIS GmbH (HIS) bzw. für die 20. Sozialerhebung das HIS Institut für Hochschulforschung (HIS-HF) als Akteur gekennzeichnet.

⁴ Da zu den Erhebungszeitpunkten der Daten keine Datennachnutzung vorgesehen war, sind einige Informationen zu den Erhebungen nicht mit dem Fokus einer späteren Datennachnutzung dokumentiert worden und deshalb teilweise nicht mehr rekonstruierbar. Dies ist an entsprechenden Stellen im Text angemerkt.

⁵ Informationen zu verfügbaren Datensätzen und Dokumentationen können im Metadaten-System des FDZ-DZHW (<https://metadata.fdz.dzhw.eu/#/de/studies/ssypool?version=1.0.1>) heruntergeladen werden.

⁶ Berücksichtigt werden hier jeweils nur die Stichproben deutscher und bildungsinländischer Studierender. Die im Rahmen der fünf einbezogenen Sozialerhebungen ebenfalls erhobenen Daten der Bildungsausländer*innen sind im Datenpaket nicht enthalten, auch deshalb, weil diese Studierenden mit einem gesonderten Fragebogen befragt worden waren.

der Variablen angewandten Prinzipien beschrieben; Kapitel 4 informiert über das Konzept der Variablennamen und Kapitel 5 enthält Informationen über die Reihenfolge und die sozialerhebungsbezogene Präsenz der Variablen im Datensatz. Im Unterschied zu den Scientific Use Files der Einzelstudien enthält der gepoolte Datensatz, der vor allem der sozialerhebungsübergreifenden Analyse dient, keine Gewichte.

Eine (weitere) Anonymisierung des gepoolten Datensatzes war nicht notwendig, weil (1) aufgrund des Querschnittsdesigns der Untersuchungsreihe mit der Datenzusammenfügung keine zusätzlichen Informationen je Fall generiert, sondern lediglich Variablen um Fälle ergänzt werden (Datensatz im long format) und (2) seine Datengrundlage die bereits faktisch anonymisierten SUF der einbezogenen Sozialerhebungen sind. Die Beschreibung der je Sozialerhebung vorgenommenen Anonymisierung sind den Daten- und Methodenberichten zu den Einzel-SUF zu entnehmen. In einigen wenigen Fällen werden anonymisierte Variablen einzelner SUF für die sozialerhebungsübergreifende Harmonisierung verwendet. Die Erstellung der Einzel-SUF erfolgte zu verschiedenen Zeitpunkten nach verschiedenen Standards (s. Tabelle 1), was auch zur Folge hatte, dass abweichende Anonymisierungskonzepte angewandt worden waren. Ein Beispiel hierfür ist das Alter der Befragten. Für die Einzel-SUF waren unterschiedliche Altersgruppen gebildet worden. Für die Harmonisierung musste zu diesem Zweck auf die anonymisierte, nicht aggregierte Altersvariable zurückgegriffen werden, um eine einheitliche Gruppierung umzusetzen. Welche Variablen darüber hinaus mittels Rückgriff auf Variablen, die im Rahmen der Einzel-SUF nicht zum Download zur Verfügung standen, harmonisiert wurden, kann anhand der Dokumentation zur Variablen-Harmonisierung nachvollzogen werden.

2 Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Daten der gepoolten Sozialerhebung werden durch das FDZ des DZHW entsprechend der europäischen Datenschutzgrundverordnung (EU-DSGVO) anonymisiert bereitgestellt und ausschließlich zur wissenschaftlichen Nutzung freigegeben.⁷ Das FDZ bietet ein *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung an.

Voraussetzungen für die Nutzung des SUF sind die Anstellung der Datennutzer*innen an einer wissenschaftlichen Einrichtung und der Abschluss eines Datennutzungsvertrags mit dem FDZ. Studierende oder Promovierende ohne eine Anstellung an einer wissenschaftlichen Einrichtung müssen gemeinsam mit einer betreuenden Person, die an einer wissenschaftlichen Einrichtung angestellt ist, einen Datennutzungsvertrag abschließen. Im Zuge des Vertragsabschlusses wird durch das FDZ das Vorliegen eines wissenschaftlichen Nutzungsinteresses geprüft. Das Formular für den Datennutzungsantrag kann von der Website des FDZ heruntergeladen werden.

⁷ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, S. 6 ff.), an dem sich bereits das Leibniz-Institut für Bildungswissenschaften (LifBi) (Koberg, 2016, S. 699 ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (Hochfellner, Müller, Schmucker und Roß, 2012, S. 9 f.) orientieren. Das FDZ des DZHW hat diesen Ansatz an die Anforderungen der eigenen Datenbestände angepasst und nutzt vier Kategorien von Maßnahmen zur Sicherstellung des Datenschutzes, die in unterschiedlicher Weise kombiniert werden können: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen.

[Datenzugang] Das SUF der gepoolten Sozialerhebung wird via Download angeboten.

- **Download:** Die Daten werden verschlüsselt auf der Website des FDZ zum Download bereitgestellt. Datennutzer(innen) können die Daten auf ihrem lokalen Computer speichern, falls gewünscht selbst mit Daten aus externen Quellen verknüpfen und die Daten mit eigener Software analysieren.

[Datenprodukte] Über den *Digital Object Identifier* (DOI) [10.21249/DZHW:ssypool:1.0.1](https://doi.org/10.21249/DZHW:ssypool:1.0.1) ist eine Website mit zentralen Informationen zur Studie, weiteren Dokumentationsmaterialien sowie einer Übersicht der zur Verfügung stehenden Datenprodukte zur Studie erreichbar.

[Gebühren der Datenbereitstellung] Das SUF wird derzeit (Stand: Juni 2019) kostenfrei zur Verfügung gestellt. Änderungen bzw. die aktuelle Gebührenordnung können auf der Website des FDZ (<https://fdz.dzhw.eu>) eingesehen werden.

[Pflichten der Datennutzer*innen] Die Datennutzer*innen sind verpflichtet, folgende Regeln⁸ einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzer(innen) Kenntnis von Sicherheitslücken hinsichtlich Datenschutz bzw. Datensicherheit erlangen, müssen diese dem FDZ unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** SUF dürfen nur durch die Person(en) genutzt werden, die den Datennutzungsvertrag abgeschlossen hat/haben.
- **Löschungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel 1,5 Jahre) von jeglichen Rechnern, Servern und Datenträgern zu löschen. Ebenso müssen alle Sicherungskopien, modifizierten Datensätze (z. B. Arbeits-, Auszugs- oder Hilfsdateien) sowie Ausdrücke vernichtet werden.
- **Bereitstellung/Meldung von Publikationen:** Jede Art von Publikation, die aus der Arbeit mit Daten des FDZ hervorgeht, ist dem FDZ unmittelbar nach Veröffentlichung anzuzeigen und – unabhängig vom Veröffentlichungsformat – als elektronische Version zur Verfügung zu stellen.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen gemäß den Vorgaben des FDZ zitiert werden.⁹

3 Harmonisierung

Die Themen der Sozialerhebung fokussieren traditionell auf die wirtschaftliche und soziale Lage der Studierenden. Anhand eines Kernkatalogs von Fragen richtete sich das Erkenntnisinteresse sowohl auf die Momentaufnahme bestimmter Parameter als auch auf ihre Entwicklung über die Zeit. Darüber hinaus waren jeweils aktuelle Themen ein- oder mehrmalig Gegenstand der Erhebung.

⁸ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

⁹ vgl. Zitation unter <https://metadata.fdz.dzhw.eu/#!/de/studies/stu-ssypool?version=1.0.1>

[Vergleichbarkeit der Messungen] Trotz der angestrebten Zeitreihen, die eine Input-Harmonisierung der Instrumente voraussetzen, weisen auch die Fragen des Kernkatalogs aus unterschiedlichen Gründen im Vergleich der fünf Sozialerhebungen vielfältige Abweichungen auf. Es können folgende Gruppen von Gründen für als Messunterschiede einzustufende Abweichungen differenziert werden:

■ **Sachbezogene Gründe:**

- Veränderungen von zu erhebenden Sachverhalten, wie z. B. Änderungen bei (Hochschul-)Abschlüssen, bei Bildungs- und Förderinstitutionen bzw. ihren Bezeichnungen, bei gesetzlichen Regelungen (z. B. Studienkredite, Studiengebühren, Wehrdienst), Herkunftsstaaten(gruppen) bzw. Zielstaaten(gruppen) für Auslandsmobilität
- Veränderungen des Erkenntnisinteresses, wie z. B. Bedarf an detaillierteren Informationen zu bestimmten Untergruppen, z. B. zu Studierenden mit Kind, ~ mit gesundheitlicher Beeinträchtigung, ~ mit besonderen Bedingungen des Hochschulzuges

■ **Methodische Gründe:**

- Anpassung als Reaktion auf das Antwortverhalten der Befragten, die z. B. das Kindergeld, das ihre Eltern erhalten, zunehmend häufiger als eigenständige Einnahmenquelle nannten oder die beim Zeitbudget angaben, dass es keine „typische“ Semesterwoche gäbe bzw. das gerade die letzte Woche „untypisch“ gewesen wäre
- Streichung von Items aus Item-Batterien, um die Länge des Fragebogens zu kürzen
- Änderungen in Zusammenhang mit dem Switch der Erhebungsmethode von einem Paper-Pencil- (17. - 20. Sozialerhebung) zu einem Online-Survey (21. Sozialerhebung). Bei dem Online-Survey erfolgte z. B. eine automatische Filterung statt der Filterung mit entsprechenden Fragen, viele Frageformulierungen waren aufgrund vorheriger Antworten individualisiert, es wurde die Möglichkeit zur Randomisierung der Abfolge von Fragen oder Items genutzt etc.

- Aufgrund des Querschnittsdesigns war eine Übereinstimmung der Variablen- und Wertetabells identischer Fragen für die Primärforschungsprojekte nicht zwingend. Diese Abweichungen wurden auch im Prozess der Erstellung der einzelnen SUF nicht harmonisiert. Die SUF-Erstellung erfolgte zudem zu verschiedenen Zeitpunkten nach unterschiedlichen Standards (vgl. Tabelle 1 und Tabelle 2), so dass auch in diesem Prozess keine Einheitlichkeit hergestellt worden war.

■ **Sonstige Gründe:**

- nicht intendierte, zufällige Abweichungen, wie z. B. geänderte Reihenfolge der Kategorien (männlich/weiblich vs. weiblich/männlich), auf Flüchtigkeit, individuellem Sprachgebrauch oder -verständnis beruhende Detailunterschiede in der Formulierung von Fragen, Items oder Antwortvorgaben.
- gesellschaftliche Veränderungen, z. B. zunehmende Akzeptanz bzw. Erwartungshaltung gegenüber der Erhebung von Kindergeld für die Befragten als eigenständige Kategorie, von eingetragenen Lebenspartnerschaften oder einer dritten Option bei der Geschlechtsangabe.

All diese Abweichungen müssen vor dem Poolen der Daten erkannt und in ihrer Systematik identifiziert werden, um für die ex post-Harmonisierung begründete Entscheidungen über (1) die Aufnahme einer Variable in den Datensatz und – im Falle ihrer Aufnahme – (2) die erforderliche bzw. mögliche Harmonisierung zu treffen.

Mit Hilfe des gepoolten SUF sollen sozialerhebungsübergreifende Analysen erleichtert werden. An diesem Ziel sind die drei nachfolgend dargestellten Hauptregeln für die Harmonisierung ausgerichtet.

[1. Vollständigkeit der Messungen] Ein zentrales Kriterium für die Aufnahme einer Variablen ist die Häufigkeit ihrer Repräsentanz in den Quell-Studien. Nur Variablen, die in mindestens drei der fünf Sozialerhebungen in vergleichbarer bzw. harmonisierbarer Weise enthalten sind, wurden in den Gesamtdatensatz aufgenommen.

[2. Referenzmessung] Der Maßstab dafür, ob eine identische oder abweichende Messung vorliegt, wurde für den vorliegenden Datensatz historisch definiert, beschränkt sich jedoch auf die einbezogenen Datensätze. Als „Urmessung“ gilt demzufolge die 17. Sozialerhebung, weil sie die älteste der fünf zu poolenden Erhebungen ist bzw. die Sozialerhebung, die eine Frage zum ersten Mal enthält. Die nachfolgenden Sozialerhebungen werden an der Formulierung einer Frage, der dazugehörigen Ausfüllanweisung, den Items sowie dem Antwortmodell der 17. Sozialerhebung bzw. der späteren Ersterhebung gemessen. Ausnahmen von dieser Regel sind aus triftigem Grund möglich. Ein Beispiel hierfür ist der angestrebte Studienabschluss. Aufgrund der 2003 noch im Umsetzungsprozess befindlichen Studienstrukturreform sind Organisation und Bezeichnung der neuen Abschlüsse Bachelor und Master im Rahmen der 17. Sozialerhebung noch von Vorläufigkeit gekennzeichnet, so dass hier auf später etablierte Abschlüsse bzw. ihre Bezeichnung zurückgegriffen wird (s. Dokumentation zur Harmonisierung).

[3. Vollständigkeit nicht-referenzielle Messungen] Sollte eine spätere Variante häufiger als die Erstmessung und mindestens dreimal eingesetzt worden sein, so wird diese übernommen und die Erstmessung nach Möglichkeit angepasst.

Ausführliche Informationen zu den vorgenommenen Harmonisierungen stellt eine entsprechende Dokumentation zum Datenpaket zur Verfügung.¹⁰

Wie oben bereits erwähnt, kommt der 21. Sozialerhebung eine Sonderstellung innerhalb der einbezogenen Datensätze zu, weil sie ausschließlich als Online-Survey durchgeführt wurde. Dieser Methodenswitch wurde – neben der Ausweitung des Fragekatalogs – auch für einen zielgruppenspezifischen Zuschnitt vieler Standardthemen genutzt, wie z.B. Zeitbudget, studentische Erwerbstätigkeit, monatliche Ausgaben. Die Formulierung der Fragen, Ausfüllanweisungen und Antwortoptionen weichen zum Teil stark ab von den methodischen Standards der Vorläuferbefragungen. Mit Blick auf das maximal mögliche Auswertungspotential des gepoolten Datensatzes für den inhaltlichen Kern der Untersuchungsreihe, wurde jeweils eine harmonisierte Variante für die Variablen der 21. Sozialerhebung entwickelt und aufgenommen. Weil es jedoch möglicherweise akzeptable, methodisch vertretbare Alternativen zur gewählten Harmonisierungs-Variante gibt bzw. diese sich u. U. nicht für jede Forschungsfrage eignet, werden neben den harmonisierten Variablen auch die jeweiligen Original-Variablen der 21. Sozialerhebung in den Datensatz aufgenommen. So können Nutzer*innen bei Bedarf eigene Adaptionen vornehmen. Welche Variablen das in Einzelnen betrifft, kann der Dokumentation zur Harmonisierung entnommen werden.

¹⁰ Middendorff, E. & Wallis, M. (2021). 17. – 21. Sozialerhebung. Dokumentation der Variablen-Harmonisierung für den gepoolten Datensatz der 17. bis 21. Sozialerhebung (2003 – 2016). DZHW: Hannover

4 Schema der Variablennamen

Das FDZ-DZHW hat einen Standard zur Variablenbenennung entwickelt, der in den hier aufbereiteten SUF und CUF angewendet wird.¹¹ Es besteht aus einer Präfix-Stamm-Suffix-Systematik: Der Variablenname enthält in Präfix und Suffix zentrale Metadaten, die für die strukturierte Verarbeitung von Variablen nötig sind. Der Stamm enthält zwei hierarchisch zusammenhängende Differenzierungen: Kennzeichnung des Themas sowie eine numerische Ordnung innerhalb des Themas.

Die systematische Vergabe von Stamm und Präfix sind unerlässlich, da sie Metadaten enthalten, die für die weitere (Meta)Datenaufbereitung notwendig sind. Nach der Evaluation der bisherigen Erfahrungen wurde die „thematische Freigabe“ des Stamms als bestes Mittel der Ressourcenverminderung bei gleichzeitig möglichst hohem Informationsgehalt identifiziert.

Im Folgenden wird das Variablennamenschema dargestellt, das für den gepoolten Datensatz der 17. – 21. Sozialerhebung entwickelt wurde und das sich am sogenannten Goldstandard des FDZ des DZHW entwickelten einheitlichen Variablennamenschema orientiert (vgl. ebenda). Die Zusammenfügung von Datensätzen setzt voraus, dass identische Variablen und/oder identische Fälle als solche eindeutig identifizierbar sind. Für den gepoolten Datensatz aus fünf Sozialerhebungen (17. – 21. Sozialerhebung) muss deshalb ein einheitliches Variablennamenschema entwickelt werden, um dieser Anforderung zu entsprechen. Darüber hinaus gibt es kohortenbezogene Variablenspezifika, wie z. B. zusätzliche Items einer Itembatte, Modifikationen in der Formulierung der Frage und/oder Antwort(en). Diese Besonderheiten sollen im Variablennamen systematisch kenntlich gemacht werden, damit Nutzer*innen sowohl die thematische Zugehörigkeit als auch die Besonderheit einer Variablen erkennen können.

4.1 Variablennamenschemata der einzubeziehenden Datensätze

Für alle fünf Quelldatensätze wurden im FDZ SUFs erstellt. Die SUF-Erstellung erfolgte zu unterschiedlichen Zeitpunkten und unter Anwendung unterschiedlicher Standards sowohl in Bezug auf die Zugangswege (Tabelle 1) als auch auf das Variablennamenschema (Tabelle 2).

Tabelle 1: Standards und Zugangswege der SUFs zur 17. – 21. Sozialerhebung

Sozialerhebung Nr.	Erhebungsjahr	Jahr der SUF-Erstellung	Angewandter SUFStandard	Zugangswege
17.	2003	2019	Bronze	SUF: Download
18.	2006	2019	Bronze	SUF: Download
19.	2009	2017	Gold	CUF: Download SUF: Download, Remote-Desktop, On-Site
20.	2012	2017	Gold	CUF: Download SUF: Download, Remote-Desktop, On-Site
21.	2016	2018	Silber	CUF: Download SUF: Download, Remote-Desktop, On-Site

¹¹ Vgl. Daniel, Weber (2018). Einheitliches Variablennamenschema für das FDZ des DZHW. Gold- und Silberstandard. Version 3.0. Projektbericht. DZHW: Hannover

Tabelle 2: Variablennamenschemata der SUFs zur 17. – 21. Sozialerhebung

Sozialerhebung Nr.	Jahr	Beispiele (Variablen-Name: Variablen-Label)	SUF-Variablenschema
17.	2003	absartagg: angestrebter Abschluss bilmut: allgemeinbildender Abschluss der Mutter baf: Im Sommersemester 2003 nach dem BAföG gefördert. ausern: Ausgaben für Ernährung	projektseitig mnemotechnisch ¹²
18.	2006	absartagg: angestrebter Abschluss bilmut: allgemeinbildender Abschluss der Mutter baf: Im Sommersemester 2006 nach dem BAföG gefördert. ausern: Ausgaben für Ernährung	projektseitig mnemotechnisch
19.	2009	stu03_g1: angestrebter Abschluss par04: allgemeinbildender Abschluss der Mutter baf01: Im Sommersemester 2009 nach dem BAföG gefördert. fin03b: Ausgaben für Ernährung	FDZ-Präfix- Stamm- Suffix- Schema
20.	2012	stu02_g1: angestrebter Abschluss par04: allgemeinbildender Abschluss der Mutter baf01: Im Sommersemester 2012 nach dem BAföG gefördert. fin03b: Ausgaben für Ernährung	FDZ-Präfix- Stamm- Suffix- Schema
21.	2016	sabsan_g1: angestrebter Abschluss deltshum: allgemeinbildender Abschluss der Mutter fbafja: Im Sommersemester 2016 nach dem BAföG gefördert. fausgernehh: Ausgaben für Ernährung, Einzel-Haushalt	projektseitiges Präfix- Stamm- Suffix- Schema

4.2 Variablennamenschema des gepoolten Datensatzes

Die Harmonisierung der Variablennamen ist zwingend erforderlich, wenn die Daten gleicher Items aus unterschiedlichen Sozialerhebungen gemerged werden sollen.

Für das Variablennamenschema des Datensatzes, der die sozialerhebungsübergreifend vergleichbaren Daten der 17. – 21. Sozialerhebung enthält, wird der im FDZ-DZHW entwickelte Goldstandard verwendet.¹³ Dieser Standard sieht vor, dass sich Variablennamen aus Präfix, Stamm und Suffix zusammensetzen. Das Präfix der Variable enthält bei Längsschnittbefragungen mit mehr als einem Befragungszeitpunkt (Panel) die Wellenkennung anhand eines Buchstabens. Da es sich bei den Sozialerhebungen um Querschnittsbefragungen handelt, entfällt das Präfix. Im Stamm geht der Themenbereich, dem die Variable zugeordnet ist, aus einem dreistelligen englischen Buchstabenkürzel hervor (Tabelle 3).

Die Unterscheidung, mit welcher der fünf Sozialerhebungen die jeweiligen Daten einer Variable erhoben wurden, erfolgt über eine entsprechende Identifikatorvariable (*ssynr*) mit den Werten 17-21.

¹² Mnemotechnischen Variablennamen werden auch als „sprechende“ Variablennamen bezeichnet. Mit ihnen vermittelt sich meist bereits der zentrale Inhalt der Variable, z. B. absart, geschl, eink, kind.

¹³ Vgl. Daniel, Weber (2017). Einheitliches Variablennamenschema für das FDZ des DZHW. Gold- und Silberstandard. Version 3.0. Projektbericht. DZHW: Hannover. S. 8 ff.

Tabelle 3: Teilelemente und Zusammensetzung des Variablenstammes

Teilelement	Beschreibung
Themen-differenzierung*	Mit einem (englischen) Kürzel aus drei Buchstaben wird die Variable einem inhaltlichen Themengebiet zugeordnet. Der gepoolte Datensatz enthält nur Themengebiete, die Bestandteil aller fünf einzubeziehenden Sozialerhebungen (17. – 21. Sozialerhebung) sind.
Nummerierung*	Innerhalb der definierten Themenbereiche werden die Variablen auf minimal zwei, maximal drei Stellen durchnummeriert.
Indizierung	Mit Hilfe eines Buchstabens am Ende des Stamms können verschiedene Variablen, die zur gleichen Frage gehören und dadurch die gleiche Themendifferenzierung und Nummerierung aufweisen (z. B. bei Itembatterien, Mehrfachnennungen oder Fragen, in denen geschlossene und offene Fragen kombiniert werden), gekennzeichnet werden (z. B. 01a, 01b, 01c, ...). Falls eine Frage den Umfang von 26 Einzelvariablen (a-z) überschreitet, wird die Itembezeichnung ab dem 27. Item mit zwei Buchstaben fortgesetzt (aa, ab, ac, ...).

* muss zwingend vergeben werden

Tabelle 4 stellt die verschiedenen Themenbereiche des gepoolten Datensatzes der 17. – 21. Sozialerhebung sowie das zugehörige Kürzel für den Stamm des Variablennamens dar. Das Kürzel leitet sich jeweils von der englischen Bezeichnung für das Themenfeld ab. Entsprechend der Zielstellung dieses Datenpools, sozialerhebungsübergreifende Analysen zu erleichtern, enthält der Datensatz ausschließlich Themenbereiche, die Gegenstand von mindestens drei der fünf einbezogenen Sozialerhebungen war und auf vergleichbare Weise erhoben wurden.

Tabelle 4: Themengebiete in den Variablennamen des gepoolten Datensatzes (17. – 21. Sozialerhebung 2003 – 2016)

Nr.	Themengebiete-Kürzel (= Stamm)	Themengebiet (englisch)	Themengebiet (deutsch)
1	dem	socio-demographic characteristics	sozio-demographische Merkmale
2	par	characteristics of parents	Merkmale der Eltern
3	stu	characteristics of study	Merkmale des Studiums
4	ped	prior education and entry into HE	Vorbildung und Hochschulzugang
5	fin	financing (of living during studies)	Finanzierung (des Lebensunterhalts während des Studiums)
6	baf	BAföG (German Federal Grant on Training and Education Promotion)	BAföG (Bundesausbildungsförderungsgesetz)
7	tim	time usage (studies/job)	Zeitaufwand (Studium/Erwerbstätigkeit)
8	job	job during studies	Erwerbstätigkeit während des Studiums
9	abr	studying abroad	studienbezogener Auslandsaufenthalt
10	lan	language skills	Sprachkenntnisse
11	liv	living (accommodation)	Wohnsituation
12	adv	demand for advice and information	Beratungs- und Informationsbedarf
13	nut	mensa and nutrition	Mensa und Ernährung

4.3 Das Suffix

Das anhand eines Unterstriches vom Stamm abgetrennte Suffix enthält verschiedene Zusatzinformationen, wie in Tabelle 5 beschrieben:¹⁴

Tabelle 5: Zusammenfassende Übersicht der verwendeten Suffix Kürzel¹⁵

Suffix-Bedeutung (Kürzel)	Beschreibung
generiert (g#)	Generierte Variablen werden mit dem Kürzel g# (g1 bzw. bei weiteren Derivaten g2, g3, ...) markiert. Unter den Typus der generierten Variablen fallen alle Variablen, die aus einer oder mehreren Variablen des Ursprungsdatensatzes erzeugt wurden (Recodierungen ¹⁶ , Indizes, vercodete Variablen, Aggregationen). Präfix und Stamm der generierten Variablen entsprechen jeweils der Ausgangsvariablen. Wird eine abgeleitete Variable aus mehreren Ausgangsvariablen gebildet, so wird der Stamm neu vergeben.
versioniert (v#)	Bei Langzeitstudien wie der Sozialerhebung können einzelne Fragen im Laufe der Zeit abgewandelt werden (neue Fragen-, Item- oder Antwortkategorienformulierung). Dabei gilt, dass mit jeglicher Formulierungsänderung (Satzstellung, Verwendung von Synonymen oder Abkürzungen) immer eine neue Version der Variable einhergeht, die entsprechend gekennzeichnet wird. Referenz für eine Änderung ist die Fassung der Frage in der jeweils ältesten der fünf einzubeziehenden Sozialerhebungen – zumeist also die 17. Sozialerhebung. Diese Variablenversionen erhalten denselben Variablenstamm wie die Ursprungsvariable, ergänzt um ein sozialerhebungsbezogenes Versionskürzel v# im Suffix (z. B. <u>_v18</u> für eine Änderung, die im Rahmen der 18. Sozialerhebung vorgenommen wurde, <u>_v19</u> für die Änderung im Rahmen der 19. Sozialerhebung usw.). Die ursprüngliche Version (hier: die Version, die als erstes in einer der fünf einzubeziehenden Sozialerhebungen eingesetzt wurde) in der Variable erhält keine Kennzeichnung. Sollte nach einer zwischenzeitlichen Änderung wieder zur Ursprungsvariante zurückgekehrt worden sein, erfolgt ebenfalls keine Kennung. Als über die Zeit „unveränderte Variable“ werden dementsprechend nur Items mit <i>exakt gleicher Formulierung</i> aller Bestandteile einer Variable (Frage, Ausfüllanweisung, Antwortmodell, inhaltlicher und/oder zeitlicher Bezug) in den gepoolten Sozialerhebungen angesehen. ¹⁷ Die Darstellung und die Anordnung der Frage im Erhebungsinstrument spielen keine Rolle. Ebenso ist nicht notwendig, dass die zugehörige Frage auch den gleichen Personengruppen wiederholt gestellt wurde.
harmonisiert (h)	Zusätzlich zu den unterschiedlichen Versionen einer Variablen können auch Variablen erzeugt werden, welche orientiert am „kleinsten gemeinsamen Nenner“ der unterschiedlichen Variablenversionen harmonisiert werden. Diese Variablen erhalten ebenfalls den Namen der Ausgangsvariablen, ergänzt um das Suffix _h .

¹⁴ Vgl. Daniel, Weber (2017). Einheitliches Variablennamenschema für das FDZ des DZHW. Gold- und Silberstandard. Version 3.0. Projektbericht. DZHW: Hannover. S. 10 f.

¹⁵ Die optionalen Suffix-Kürzel **p** für Variablen, die im Zuge des Datenaufbereitungsprozesses plausibilisiert wurden bzw. das Suffix-Kürzel für Varianten des Zugangsweges (c, d, r, o, a) entfallen, weil bereits aufbereitete SUF gepoolt werden und nur ein Zugangsweg (Download) zur Verfügung gestellt wird.

¹⁶ Darauf hinzuweisen ist, dass Variablen, bei denen lediglich im Rahmen der Plausibilisierung Editionen vorgenommen wurden, nicht als generierte Variablen anzusehen sind.

¹⁷ Auf Zeitvergleiche angelegte Langzeituntersuchungsreihen sollten gewährleisten, dass Änderungen an Fragetexten und/oder Antwortkategorien von Variablen nur dann erfolgen, wenn dies inhaltlich und/oder methodisch zwingend geboten ist. (Vermieden werden sollte vor allem das Einfügen oder Löschen von Füllwörtern und das Ersetzen von Wörtern durch Synonyme.)

4.4 Systematik fehlender Werte

Die Quell-Datensätzen weisen verschiedene Missing-Systematiken auf (vgl. Tabelle 6), die für den gepoolten Datensatz übernommen und lediglich mit einer einheitlichen Kodierung versehen wurden (vgl. Tabelle 7). Eine Harmonisierung der Missings ist nur begrenzt möglich, weil nicht alle Missings gleichermaßen existent sind bzw. vergeben worden waren. So weist die 21. Sozialerhebung online-survey-spezifische Missings auf („vorheriger Befragungsabbruch“, „nicht gesehen“). Das Missing „nicht gesehen“ wurde hier seitens des Primärforschungsprojektes auch vergeben, wenn Befragte aufgrund ihres vorherigen Antwortverhaltens über Fragen hinweggefiltert wurden. Bei der 17. – 20. Sozialerhebung enthalten gefilterte Fragen das entsprechende Missing „filterbedingt fehlend“; bei der 21. Sozialerhebung hingegen werden diese Missings unter „nicht gesehen“ subsummiert¹⁸. Da der gepoolte Datensatz auf der Grundlage der fünf Einzel-SUF erstellt wird und in dem SUF für die 21. Sozialerhebung diese pauschalisierende Kodierung des Primärforschungsprojektes unverändert blieb, wurde auf eine ex post Plausibilisierung des verschiedene Gründe für fehlende Werte zusammenfassenden Missings verzichtet.

Tabelle 6: Systematik fehlender Werte in den Quell-Datensätzen (17. – 21. Sozialerhebung)

17. u. 18. Sozialerhebung		19. u. 20. Sozialerhebung		21. Sozialerhebung	
Code	Wertelabel	Code	Wertelabel	Code	Wertelabel
-2	keine Angabe	-999	weiß nicht	-12	weiß nicht (selbstberichtet)
		-998	keine Angabe	-9990	nicht beantwortet
		-997	keine Angabe (Antwortkategorie)	-13	keine Angabe (selbstberichtet)
		-996	Interviewabbruch	9993	-vorheriger Befragungsabbruch
		-994	verweigert		
-1	trifft nicht zu	-989	filterbedingt fehlend	-9981	Filter-Plausitz (selbstberichtet)
		-988	trifft nicht zu	-11	tnz (selbstberichtet)
		-987	designbedingt fehlend (Fragebogensplit)	-9991	nicht gesehen (Filter, Split, Einblendbedingung)
		-985	designbedingt fehlend (Kohorte) ^b		
		-969	unbekannter fehlender Wert	-9982	inhaltl. Plausitz
		-968	unplausibler Wert ^d		
		-967	anonymisiert	-967	anonymisiert
		-966	nicht bestimmbar ^e	-9996	Zusatzvariable nicht bestimmbar
		-965	ungültige Mehrfachnennung		

¹⁸ Beispiel hierfür ist die Variable „Anzahl Kinder“ (dem06_h): Für die 21. Sozialerhebung werden fast 53.000 „nicht gesehen“-Fälle ausgewiesen, von denen mehr als 52.000 zuvor die Frage, ob sie Kinder haben, verneint hatten.

Ein weiteres Beispiel für die Restriktionen bei der Missing-Harmonisierung ist die Erhebung des Geschlechts: Die 21. Sozialerhebung enthielt erstmals eine dritte Antwortoption. Diese wurde für den SUF anonymisiert, indem sie zusammen mit „keine Angabe“ als -9996 „Zusatzvariable nicht bestimmbar“ kategorisiert wurde. Einerseits können und sollen die tatsächlichen „k. A.“-Fälle der 21. Sozialerhebung nicht mehr getrennt von jenen werden, die die dritte Antwortvariante genutzt hatten, andererseits wäre es falsch, die tatsächlichen „k. A.“ der übrigen vier Sozialerhebungen auf „Zusatzvariable nicht bestimmbar“ zu setzen.

Für den gepoolten Datensatz wird die nachfolgende Missings-Systematik umgesetzt (vgl. Tabelle 7). Eine besondere Bedeutung kommt dem Code -985 „designbedingt fehlend“ zu: Er wird für Variablen vergeben, die zwar in den gepoolten Datensatz aufgenommen wurden, aber in einzelnen Quell-Datensätzen nicht oder in nicht harmonisierbarer Weise vorhanden sind. Es wird ein zusätzliches Missing eingeführt: -986 „ssy-spezifisch fehlend“. Es kennzeichnet fehlende Werte einer Variable, die deshalb fehlen, weil bei einer oder mehreren Sozialerhebungen abweichende Varianten der Frage zu Einsatz kamen, die in einer spezifischen, suffix-gekennzeichneten Variable abgelegt wurden. Sollte eine Variable oder ein Wert in einer oder mehreren Sozialerhebungen gänzlich fehlen, wird das mit den Missing -985 „kohortenspezifisch fehlend“, wobei die einzelnen Sozialerhebungen jeweils mit einer Kohorte gleichgesetzt werden.

Tabelle 7: Systematik fehlender Werte im gepoolten Datensatz der 17. – 21. Sozialerhebung

Wertebereich	Code	Wertelabel
(selbstberichteter) Nonresponse	-999	weiß nicht (Antwortkategorie)
	-998	keine Angabe
	-997	keine Angabe (Antwortkategorie)
Nicht zutreffend / filterbedingt fehlend	-989	filterbedingt fehlend
	-988	trifft nicht zu
Designbedingt fehlend	-987	designbedingt fehlend (Split, Filter, Einblendbedingung) ^a
	-986	ssy-spezifisch fehlend ^b
	-985	kohortenspezifisch fehlend ^c
Editierter fehlender Wert	-969	unbekannter fehlender Wert ^d
	-968	unplausibler Wert ^e
	-967	anonymisiert
	-966	nicht bestimmbar ^f
	-965	ungültige Mehrfachnennung

a Dieses Missing kommt nur in der 21. Sozialerhebung vor. Sie wurde als einzige online durchgeführt und hat Splits und Einblendbedingungen eingesetzt. Zu den Einblendbedingungen gehören auch Filterführungen, in deren Ergebnis Folgefragen nicht angezeigt wurden, d. h. von den Befragten „nicht gesehen“ wurden – ebenso wie bei Splits.

b Dieser Wert wird vergeben, wenn eine Variable ein Missing aufweist, weil es eine sozialerhebungsspezifische Variante dieser Variablen gibt. Diese Varianten sind in einer eigenen Variable im Datensatz enthalten und mit einem entsprechenden Suffix im Variablennamen gekennzeichnet.

c Dieses Missing bedeutet, dass die Variable bzw. der Wert in einer oder mehreren Sozialerhebungen nicht erhoben wurde/nicht enthalten ist.

d Dieser Missing wird vergeben, wenn keinerlei Ursache rekonstruiert werden kann.

e Angaben, die aufgrund unterschiedlicher Faktoren in der Codierphase als nicht plausibel eingestuft werden, erhalten diesen Wert. Eine exakte Rekonstruktion ist ggf. nicht mehr möglich.

f Diese Kategorie wird vergeben, wenn eine eindeutige Codierung nicht möglich ist, z. B. offene Angabe, die nicht vercodet werden konnte, da sie nicht lesbar ist.

5 Variablen im Datensatz

5.1 Sortierung der Variablen im Datensatz

Da die Reihenfolge der Fragen des Fragekatalogs im Vergleich der einbezogenen Sozialerhebungen unterschiedlich ist, sind die Variablen im Datensatz thematisch gruppiert (s. Kapitel 5.2).

Generierte Variablen werden, wenn möglich, unterhalb der Ausgangsvariablen positioniert. Wurde eine neue Variable aus verschiedenen Ausgangsvariablen generiert, wird sie hinter jene Variable(ngruppe) einsortiert, welche ihr thematisch am ehesten entspricht. Falls eine thematische Zuordnung nicht möglich ist, wird die Variable an das Ende des Datensatzes gestellt.

5.2 Übersicht zur sozialerhebungsbezogenen Variablen-Präsenz

Das Scientific Use File der gepoolten Daten enthält 564 Variablen, darunter drei Systemvariablen (Ident-Nr., Nr. der Sozialerhebung und Jahr der Sozialerhebung). Die einbezogenen Sozialerhebungen sind in den 561 inhaltlichen Variablen unterschiedlich stark repräsentiert (vgl. Tabelle 8), was sowohl mit der Länge als auch mit der Spezifik des jeweiligen Fragekatalogs zusammenhängt. Der Datensatz enthält neben harmonisierten Variablen ($n = 80$) eine Vielzahl an sozialerhebungsspezifischen Variablen. Letztere wurden aufgenommen, um den Datennutzer*innen die Möglichkeit einzuräumen, je nach Forschungsfrage eigene Anpassungen vorzunehmen bzw. zu entscheiden, ob die hiermit gekennzeichneten methodischen Abweichungen für ihre Analysezwecke relevant sind.

Der große Anteil an Variablen mit dem Suffix *_v21 (40,5 %, vgl. Tabelle 8) hängt damit zusammen, dass die erstmals als Online-Survey erhobene 21. Sozialerhebung die Zielgruppenspezifika des Fragekatalogs in den Vordergrund stellen konnte und gestellt hat. So wurden viele Kernfragen an Untergruppen angepasst, was die Vergleichbarkeit und das Harmonisierungspotential stark einschränkt. Das betrifft vor allem Fragen zu den monatlichen Ausgaben, die differenziert für vier Haushaltstypen gestellt wurden, und Fragen nach dem Zeitbudget, bei dem sieben Fallgruppen unterschieden wurden (s. auch Dokumentation zur Variablen-Harmonisierung von Middendorff/Wallis).

Ein Indikator für das Potential des Datensatzes für Zeitreihenanalysen ist die Repräsentanz der einbezogenen Kohorten in den Variablen. Ein realistisches Bild lässt sich zeichnen, wenn die Variablen, die spezifisch für die 21. Sozialerhebung sind *_v21, unberücksichtigt bleiben: Nicht ganz ein Drittel aller Variablen enthält Werte für alle fünf Erhebungszeiträume (30,8 %). Jeweils deutlich mehr als ein Fünftel der Variablen ermöglichen einen Vergleich von vier bzw. drei Erhebungssemestern (21,9 % bzw. 21,6%). Ein knappes Fünftel der Variablen betreffen ausschließlich eine einzelne Sozialerhebung (19,8 %) – würden die *_v21-Variablen einbezogen, beträfe das sogar deutlich mehr als die Hälfte aller Variablen im Datensatz (52,2 %).

Tabelle 8: Übersicht zur sozialerhebungsbezogenen Variablen-Präsenz

Variablen je Sozialerhebung	Sozialerhebung Nr.					Gesamt
	17	18	19	20	21	
Anzahl	222	257	263	257	357	561
Anteil an allen Variablen im Datensatz in %	39,6	45,8	46,9	45,8	63,6	100
Variablen mit Suffix						
*_h = harmonisiert, *_# = sozialerhebungsspezifisch	Suffix					Gesamt
	*_h	*_v18	*_v19	*_v20	*_v21	
Anzahl	80	27	6	31	227	371
Anteil an allen Variablen im Datensatz in %	14,3	4,8	1,1	5,5	40,5	66,1
Variablenbezogene Präsenz der Sozialerhebungen – Zeitreihenpotential						
	Anzahl Sozialerhebungen					Gesamt
	1	2	3	4	5	
Anzahl	293	20	72	73	103	561
Anteil an allen Variablen im Datensatz in %	52,2	2,3	13,9	13,0	18,5	100
Abzügl. *_21-Variablen (n=227): Variablenbezogene Präsenz der Sozialerhebungen – Zeitreihenpotential						
	Anzahl Sozialerhebungen					Gesamt
	1	2	3	4	5	
Anzahl	66	20	72	73	103	334
Anteil an allen übrigen Variablen im Datensatz in %	19,8	6,0	21,6	21,9	30,8	100,0

Eine Übersichtstabelle über alle Variablen des gepoolten Datensatzes der 17. – 21. Sozialerhebung einschließlich der Kennzeichnung, aus welchen Erhebungen sie jeweils zur Verfügung stehen, ist Bestandteil der Dokumentation der Variablen-Harmonisierung (Middendorff/Wallis, 2021).

Literaturverzeichnis

- Becker, K., Baillet, F. & Weber, A. (2019). 21. Sozialerhebung. Daten- und Methodenbericht zur Studierendenbefragung 2016. Hannover: FDZ-DZHW. [https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy21\\$-2.0.0/attachments/ssy21_MethodReport_de.pdf](https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy21$-2.0.0/attachments/ssy21_MethodReport_de.pdf)
- Daniel, A. & Weber, A. (2017). Einheitliches Variablennamenschema für das FDZ des DZHW. Gold- und Silberstandard. Version 3.0. Projektbericht. Hannover: FDZ-DZHW.
- Daniel, A., Sarcletti, A. & Vietgen, S. (2017). 20. Sozialerhebung. Daten- und Methodenbericht zur Studierendenbefragung 2012. Hannover: FDZ-DZHW. [https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy20\\$-1.0.0/attachments/ssy20_MethodReport_de.pdf](https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy20$-1.0.0/attachments/ssy20_MethodReport_de.pdf)
- Ebel, T. & Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten (Forschungsdaten Bildung informiert Nr. 3). Verbund Forschungsdaten Bildung.
- Hoffstätter, U. & Sarcletti, A. (2017). 19. Sozialerhebung. Daten- und Methodenbericht zur Studierendenbefragung 2009. Hannover: FDZ-DZHW. [https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy19\\$-1.0.0/attachments/ssy19_MethodReport_de.pdf](https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy19$-1.0.0/attachments/ssy19_MethodReport_de.pdf)
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer & J. Skopec (Hrsg.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (S. 691–708). Wiesbaden: Springer VS. doi:10.1007/978-3-658-11994-2
- Middendorff, E. & Wallis, M. (2021): 17. - 21. Sozialerhebung. Dokumentation der Variablen-Harmonisierung und sozialerhebungsbezogene Übersicht aller Variablen für den gepoolten Datensatz der 17. bis 21. Sozialerhebung (2003 - 2016). Hannover: FDZ-DZHW. [https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-ssypool\\$attachments/Doku_Harmonisierung_Var_Uebersicht.pdf](https://metadata.fdz.dzhw.eu/public/files/data-packages/stu-ssypool$attachments/Doku_Harmonisierung_Var_Uebersicht.pdf)
- Middendorff, E. & Hoffstätter, U.: (2020). 17. Sozialerhebung. Daten- und Methodenbericht zur Studierendenbefragung 2003. Hannover: FDZ-DZHW. [https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy17\\$-2.0.0/attachments/ssy17_MethodReport_de.pdf](https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy17$-2.0.0/attachments/ssy17_MethodReport_de.pdf)
- Middendorff, E. (2019). Die Sozialerhebungen des Deutschen Studenten-werks 1951 – 2016. Ein historischer Überblick über Akteure, Wellen, Methoden, Themen und projektbezogene Publikationen (Working Paper). Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). http://www.sozialerhebung.de/download/geschichte/Geschichte_Sozialerhebung_1-21.pdf
- Middendorff, E. & Hoffstätter, U.: (2019). 18. Sozialerhebung. Daten- und Methodenbericht zur Studierendenbefragung 2006. Hannover: FDZ-DZHW. [https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy18\\$-1.0.0/attachments/ssy18_MethodReport_de.pdf](https://metadata.fdz.dzhw.eu/public/files/studies/stu-ssy18$-1.0.0/attachments/ssy18_MethodReport_de.pdf)
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(2), 581–592.