

Andreas Daniel / Andreas Sarcletti / Sandra Vietgen

# 20. Sozialerhebung

Daten- und Methodenbericht zur Studierendenbefragung  
2012

Version 1.0.0

## Daten- und Methodenbericht

Juni 2017

Autor(inn)en:  
Andreas Daniel  
Dr. Andreas Sarcletti  
Sandra Vietgen

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

Geschäftsführung:  
Prof. Dr. Monika Jungbauer-Gans  
Dr. Bernhard Hartung

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

Abbildungsverzeichnis .....	IV
Tabellenverzeichnis .....	IV
I Einleitung .....	1
II Übersicht zur 20. Sozialerhebung (2012) .....	2
III Datennutzungshinweise .....	4
1 Inhalt und Anlage der Studie .....	8
2 Erhebungsinstrument .....	11
2.1 Inhalte der Erhebungsinstrumente .....	11
2.2 Pretest .....	12
3 Grundgesamtheit und Stichprobenverfahren .....	14
4 Durchführung der Erhebung .....	15
5 Rücklauf .....	16
6 Datenaufbereitung .....	17
6.1 Datenübertragung .....	17
6.2 Codierung offener Angaben .....	18
6.3 Datenprüfung und Datenbereinigung .....	20
6.4 Generierung von Variablen .....	22
6.5 Erstellung des Datensatzes .....	22
6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels .....	22
6.7 Codierung fehlender Werte .....	23
7 Gewichtung .....	26
7.1 Vorgehen und Anwendungshinweise .....	26
7.2 Gewichtung des Datensatzes .....	27
8 Anonymisierung .....	29
9 Literaturverzeichnis .....	35

## Abbildungsverzeichnis

Abbildung 1: Datenzugangswege und Analysepotential.....	5
Abbildung 2: Rücklauf der 20. Sozialerhebung (2012) im Zeitverlauf.....	17
Abbildung 3: Datenzugangsweg, statistischer Anonymisierungsgrad und Analysepotential der Daten der 20. Sozialerhebung (2012).....	30

## Tabellenverzeichnis

Tabelle 1: Brutto- und Nettostichprobe sowie Rücklaufquote der 20. Sozialerhebung (2012).....	16
Tabelle 2: Vercodete Merkmale und verwendete Codierschemata in der 20. Sozialerhebung (2012).....	20
Tabelle 3: Themengebiete im Variablennamen der 20. Sozialerhebung (2012).....	23
Tabelle 4: Systematik fehlender Werte des FDZ-DZHW.....	25
Tabelle 5: Bereitgestellte Gewichte zur 20. Sozialerhebung (2012).....	27
Tabelle 6: Maßnahmen der statistischen Anonymisierung der Daten der 20. Sozialerhebung (2012) nach Zugangsweg.....	32

## I Einleitung

Die Sozialerhebung des Deutschen Studentenwerks (DSW) ist eine seit 1951 bestehende Untersuchungsreihe zur wirtschaftlichen und sozialen Situation der Studierenden in Deutschland.<sup>1</sup> Sie wird seit 1982<sup>2</sup> (10. Sozialerhebung) im Auftrag des DSW durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)<sup>3</sup> durchgeführt. Seit der 6. Sozialerhebung (1967/1968) wird die Studie durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert (Middendorff 2016: 2). Die Sozialerhebung dient – in Ergänzung zur amtlichen Hochschulstatistik – dem nationalen Bildungsmonitoring.

Im Rahmen des vom BMBF geförderten Projektes zum Aufbau eines Forschungsdatenzentrums für Hochschul- und Wissenschaftsforschung am DZHW (FDZ-DZHW) werden die Daten einiger jüngerer Erhebungen der Reihe nachträglich zum Zweck der Nachnutzung aufbereitet und dokumentiert. Sie werden über verschiedene Zugangswege als *Scientific Use Files* (SUF) für die wissenschaftliche Sekundärnutzung bzw. als *Campus Use Files* (CUF) für Lehr- und Übungszwecke zur Verfügung gestellt. Neben den Datensätzen der Erhebungen werden auch Dokumentationsmaterialien zu den Datensätzen und zur Durchführung der Studien bereitgestellt.<sup>4</sup>

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zur 20. Sozialerhebung (doi: 10.21249/DZHW:ssy20:1.0.0).<sup>5</sup> Weitere Dokumentationsmaterialien zur Studie (Datensatzreport, Fragebogen, Filterführungsdiagramm etc.) können frei im Metadaten-suchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden.

Abschnitt II dieses Berichts stellt die Eckdaten der 20. Sozialerhebung im Überblick dar. Die zentralen Informationen zur Nutzung der Daten dieser Studie folgen in Abschnitt III. Kapitel 1 stellt Inhalt und Anlage der Sozialerhebung bis 2012<sup>6</sup> im Allgemeinen und der 20. Sozialerhebung im Speziellen vor. Die weitere Gliederung des Berichts orientiert sich im Wesentlichen am Ablauf des Forschungsprozesses. In Kapitel 2 wird das Erhebungsinstrument beschrieben. In den Kapiteln 3 bis 6 werden der Erhebungsprozess (Stichprobenziehung, Erhebungsablauf, Rücklauf) und die Datenaufbereitung dargestellt. In den Kapiteln 7 und 8 folgt die Beschreibung der vorgenommenen Gewichtung und Anonymisierung.

<sup>1</sup> Für aktuelle Informationen zur Sozialerhebung siehe die Website des Projekts (<http://www.sozialerhebung.de>).

<sup>2</sup> Die 1. (1951) und 2. Sozialerhebung (1953) wurde jeweils vom Verband Deutscher Studentenwerke durch Gerhard Kath durchgeführt. Die 3. (1956) bis 9. Sozialerhebung (1979) wurde vom Deutschen Studentenwerk (DSW) ebenfalls durch Gerhard Kath (Geschäftsführer des Studentenwerks Frankfurt am Main) durchgeführt. Bis zur 5. Sozialerhebung (1963) wurde die Sozialerhebung vom Bundesministerium des Innern finanziell gefördert.

<sup>3</sup> 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 2013 durchgeführt wurde. Auf allen Originaldokumenten der 20. Sozialerhebung (Fragebogen, Flyer etc.) sowie in den Berichten zum Projekt ist entsprechend das HIS-Institut für Hochschulforschung (HIS-HF) als Akteur gekennzeichnet.

<sup>4</sup> Informationen zu verfügbaren Datensätzen und Dokumentationen können im Metadaten-suchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu/>) heruntergeladen werden.

<sup>5</sup> Berücksichtigt wird hier nur die Stichprobe deutscher Studierender und Bildungsinländer(innen), die mittels eines Papierfragebogens befragt wurden. Die im Rahmen der 19. Sozialerhebung ebenfalls erhobenen Daten der Bildungsausländer(innen) sowie eine online erhobene Teilstichprobe werden hier nicht dokumentiert (vgl. Kap. 1).

<sup>6</sup> Es werden nur die Sozialerhebungen bis zur hier dokumentierten 20. Sozialerhebung (2012) berücksichtigt. Die 21. Sozialerhebung (2016) wird ebenfalls zur Datennachnutzung aufbereitet werden. Für die 19. Sozialerhebung stehen bereits ein Scientific-Use-File (Datenzugang über On-Site, Remote-Desktop und Download) und ein Campus Use File zur Verfügung.

## II Übersicht zur 20. Sozialerhebung (2012)

<b>Studienreihe</b>	Sozialerhebung des Deutschen Studentenwerks (DSW)
<b>Erhebung</b>	20. Sozialerhebung (2012)
<b>Erhebende Institution</b>	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)
<b>Förderung</b>	Bundesministerium für Bildung und Forschung (BMBF)
<b>Projektmitarbeiter(innen) (Projektleitung)</b>	Beate Apolinarski, Maren Kandulla, <u>Elke Middendorff</u> , Nicolai Netz, Jonas Poskowsky
<b>Themen</b>	Angaben zum Studium Zeitbudget Finanzierung des Studiums und Erwerbstätigkeit Wohnsituation Gesundheitliche Beeinträchtigung Studienbezogene Auslandserfahrungen
<b>Erhebungsdesign</b>	Querschnitterhebung (Trend-Design)
<b>Grundgesamtheit</b>	Studierende mit deutscher Staatsangehörigkeit und Bildungsinländer(innen), die im Sommersemester 2012 an einer staatlichen oder staatlich anerkannten Hochschule in der Bundesrepublik Deutschland eingeschrieben waren (mit Ausnahme der Verwaltungsfachhochschulen, Hochschulen des Fernstudiums und Universitäten der Bundeswehr)
<b>Stichprobe</b>	Einfach (disproportional) geschichtete Zufallsstichprobe
<b>Erhebungsmethode</b>	Standardisierte postalische Befragung
<b>Erhebungszeitraum</b>	29. Mai bis 20. August 2012
<b>Auswertbare Fälle</b>	n = 15.128
<b>Rücklaufquote</b>	28,0%
<b>Datenprodukte und Zugangswege</b>	CUF: Download SUF: Download, Remote-Desktop, On-Site
<b>DOI</b>	10.21249/DZHW:ssy20:1.0.0
<b>Weitere Informationen</b>	<a href="https://fdz.dzhw.eu">https://fdz.dzhw.eu</a>

**Projektpublikationen**

Apolinarski, B. (2014). *Die wirtschaftliche und soziale Lage der Studierenden in Köln 2012. Regionalauswertung der 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung* (Kölner Studentenwerk, Hrsg.), Köln. Verfügbar unter [http://kstw.de/images/stories/presse/regionalbericht\\_kstw\\_12%2002%2014\\_fin.pdf](http://kstw.de/images/stories/presse/regionalbericht_kstw_12%2002%2014_fin.pdf)

Kandulla, M. (2014). *Die wirtschaftliche und soziale Lage der Studierenden in Nordrhein-Westfalen. Sonderauswertung der Arbeitsgemeinschaft der Studentenwerke NRW zur 20. Sozialerhebung des Deutschen Studentenwerks im Jahre 2012* (Arbeitsgemeinschaft der Studentenwerke NRW, Hrsg.), Bielefeld. Verfügbar unter [http://www.studentenwerke-nrw.de/CMS/images/stories/STWeData/documents/PDFs/Regionalbericht\\_ARGE\\_STWNRW\\_Gesamtbericht\\_06\\_02\\_14.pdf](http://www.studentenwerke-nrw.de/CMS/images/stories/STWeData/documents/PDFs/Regionalbericht_ARGE_STWNRW_Gesamtbericht_06_02_14.pdf)

Middendorff, E. & Poskowsky, J. (2014). *Die wirtschaftliche und soziale Lage der Studierenden in Hannover. Sonderauswertung der 20. Sozialerhebung des Deutschen Studentenwerks 2012* (Studentenwerk Hannover, Hrsg.), Hannover. Verfügbar unter <http://www.studentenwerk-hannover.de/fileadmin/daten/pdf/allgemein/StwH-Sozialerhebung-2012.pdf>

Apolinarski, B. (2013). *Die wirtschaftliche und soziale Lage der Studierenden in München. Regionalauswertung der 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung* (Studentenwerk München, Hrsg.). München: München. Verfügbar unter [http://www.studentenwerk-muenchen.de/fileadmin/studentenwerk-muenchen/publikationen/broschueren/stwm\\_Sozialbericht\\_20140326\\_web.pdf](http://www.studentenwerk-muenchen.de/fileadmin/studentenwerk-muenchen/publikationen/broschueren/stwm_Sozialbericht_20140326_web.pdf)

Apolinarski, B. & Poskowsky, J. (2013). *Ausländische Studierende in Deutschland 2012. Ergebnisse der 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt vom Deutschen Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) (BMBF, Hrsg.)*, Bonn. Verfügbar unter [http://www.sozialerhebung.de/download/20/soz20\\_auslaenderbericht.pdf](http://www.sozialerhebung.de/download/20/soz20_auslaenderbericht.pdf)

Kandulla, M. (2013). *Die wirtschaftliche und soziale Lage der Studierenden in Berlin. Regionalauswertung Berlin der 20. Sozialerhebung des Deutschen Studentenwerks*. (Studentenwerk Berlin., Hrsg.), München. Verfügbar unter <http://www.studentenwerk-berlin.de/studentenwerk/dokumente/65%20%7C%20Soziale%20Lage%20Studierende.pdf>

Middendorff, E., Apolinarski, B., Poskowsky, J., Kandulla, M. & Netz, N. (2013). *Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2012. 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch das HIS-Institut für Hochschulforschung* (BMBF, Hrsg.), Bonn. Verfügbar unter [http://www.sozialerhebung.de/download/20/soz20\\_hauptbericht\\_gesamt.pdf](http://www.sozialerhebung.de/download/20/soz20_hauptbericht_gesamt.pdf)

**Publikationen zum Datensatz (Auswahl)**

Staneva, M. *Studieren und Arbeiten. Die Bedeutung der studentischen Erwerbstätigkeit für den Studienerfolg und den Übergang in den Arbeitsmarkt*. (DIW Roundup Nr. 70). Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).

Netz, N. (2014). Der Zugang zu studienbezogenen Auslandsaufenthalten aus Sicht von Hochschulpolitik und Hochschulforschung: Eine Bestandsaufnahme. In U. Banscherus, M. Bülow-Schramm, K. Himpele, S. Staack & S. Winter (Hrsg.), *Übergänge im Spannungsfeld von Expansion und Exklusion. Eine Analyse der Schnittstellen im deutschen Hochschulsystem* (GEW Materialien aus Hochschule und Forschung, Bd. 121, S. 81–97). Bielefeld: Bertelsmann. doi:10.3278/6001596w

### III Datennutzungshinweise

**[Voraussetzungen der Datennutzung]** Die Daten der 20. Sozialerhebung werden durch das FDZ-DZHW entsprechend dem Bundesdatenschutzgesetz (vgl. § 40 Abs. 1 und Abs. 2 BDSG) ausschließlich zur wissenschaftlichen Nutzung und anonymisiert bereitgestellt.<sup>7</sup> Das FDZ bietet ein *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung und ein *Campus Use File* (CUF) für Lehr- und Übungszwecke an.

Voraussetzungen für die Nutzung des SUF sind die Anstellung der Datennutzerin/des Datennutzers an einer wissenschaftlichen Einrichtung und der Abschluss eines Datennutzungsvertrags mit dem FDZ. Studierende oder Promovierende ohne eine Anstellung müssen eine Zusammenarbeit mit einer/einem betreuenden Mitarbeiter(in) einer wissenschaftlichen Einrichtung nachweisen. Im Zuge des Vertragsabschlusses wird durch das FDZ auch das Vorliegen eines wissenschaftlichen Nutzungsinteresses geprüft. Das Formular für den Datennutzungsantrag kann von der Website des FDZ heruntergeladen werden. Für die Nutzung des CUF sind lediglich Name und Nutzungszweck anzugeben. Danach wird das CUF durch das FDZ-DZHW übermittelt.

**[Datenzugang]** Das CUF der 20. Sozialerhebung kann nach Bereitstellung am lokalen Computer genutzt werden. Das SUF wird über drei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind.

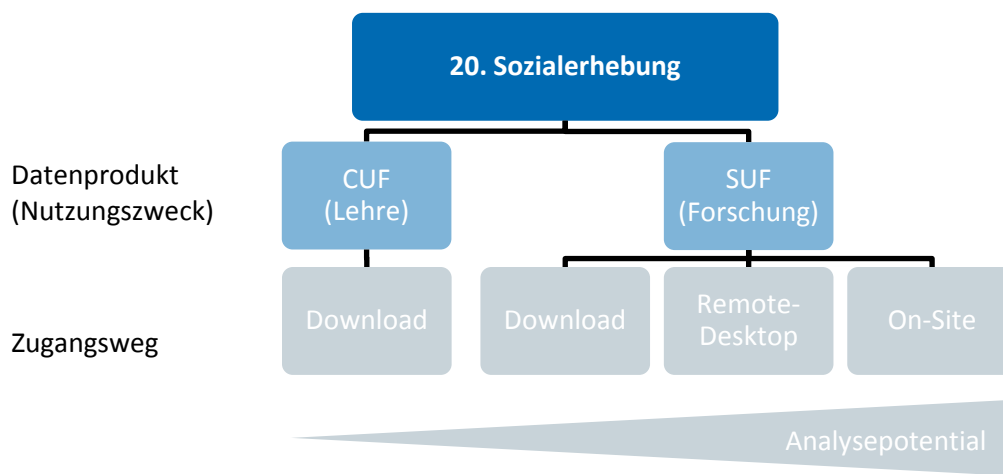
- **Download:** Die Daten werden verschlüsselt per E-Mail versandt oder auf der Website des FDZ zum Download bereitgestellt. Datennutzer(innen) können die Daten auf ihrem lokalen Computer speichern, falls gewünscht selbst mit Daten aus externen Quellen verknüpfen und die Daten mit eigener Software analysieren.
- **Remote-Desktop:** Die Daten werden auf einem Terminal-Server des FDZ-DZHW bereitgestellt. Über eine besonders gesicherte Verbindung zwischen dem lokalen Computer der nutzenden Person und dem Terminal-Server des FDZ können die Daten mit der auf dem Terminal-Server vorhandenen Software analysiert werden. Das Übertragen der Daten auf den lokalen Computer ist nicht möglich. Analyseergebnisse werden erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.
- **On-Site:** Die Daten werden in den Räumlichkeiten des FDZ-DZHW in einer kontrollierten Umgebung an einem speziell gesicherten Computer zur Analyse bereitgestellt. Wie beim Remote-Desktop-Zugang werden Analyseergebnisse erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.

<sup>7</sup> 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 Bildungsverläufe (LifBi) (vgl. Koberg, 2016, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker und Roß, 2012, S. 9 f.) orientieren. Das FDZ-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.



Die bereitgestellten Daten weisen je nach Zugangsweg einen unterschiedlich hohen Informationsgehalt auf und unterscheiden sich damit hinsichtlich ihres Analysepotentials (vgl. Abbildung 1). Dabei gilt: Je stärker der Datenzugriff der Nutzer(innen) durch technische und organisatorische Maßnahmen kontrolliert wird, desto mehr Informationen können für die Datennutzer(innen) bereitgestellt werden.<sup>8</sup> Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

**Abbildung 1: Datenzugangswege und Analysepotential**



**[Datenprodukte]** Über den *Digital Object Identifier* (DOI) 10.21249/DZHW:ssy20:1.0.0 ist eine Website mit zentralen Informationen zur Studie, weiteren Dokumentationsmaterialien sowie einer Übersicht der zur Verfügung stehenden Datenprodukte zur Studie erreichbar.

Die Daten der 20. Sozialerhebung werden über jeden im FDZ-DZHW angebotenen Zugangsweg – jeweils mit zugangswegspezifischem Analysepotential (vgl. Abbildung 1) – bereitgestellt. Das Download-CUF und das Download-SUF sind jeweils im Stata- und im SPSS-Format verfügbar. Für die Zugangswege Remote-Desktop und On-Site werden die Datensätze standardmäßig im Stata-Format bereitgestellt.

**[Gebühren der Datenbereitstellung]** CUF und SUF werden derzeit (Stand: Juni 2017) 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 Regeln<sup>9</sup> einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.

<sup>8</sup> Zu den verschiedenen Anonymisierungsgraden und Analysepotentialen des CUF und der verschiedenen SUF-Varianten vgl. Kapitel 8.

<sup>9</sup> Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

- **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-DZHW unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** SUF dürfen nur durch die Person genutzt werden, die den Datennutzungsvertrag abgeschlossen hat. CUFs dürfen ausschließlich im Rahmen der angegebenen Lehrveranstaltung weitergegeben werden.
- **Löschungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel drei 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-DZHW hervorgeht, muss dem FDZ im Voraus gemeldet und nach Veröffentlichung unverzüglich in elektronischer Form zur Verfügung gestellt werden. Informationen zu bereits vorhandenen Veröffentlichungen können dem Metadatensuchsystem<sup>10</sup> entnommen werden.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut folgender Vorgaben zitiert werden.

#### [Zitation]

- **Verwendeter Datensatz:**  
Middendorff, E., Apolinarski, B., Poskowsky, J., Kandulla, M. & Netz, N. (2013). 20. Sozialerhebung. Aufbereitet durch Daniel, A., Sarcletti, A. & Vietgen, S., doi: 10.21249/DZHW:ssy20:1.0.0, DATENSATZNAME<sup>11</sup>, released 2017. Hannover: FDZ-DZHW.

---

<sup>10</sup> <https://metadata.fdz.dzhw.eu>

<sup>11</sup> An dieser Stelle bitte den genauen Namen der verwendeten Datensatzversion angeben, z. B. ssy20\_d\_1-0-0.dta für das Download-SUF der 20. Sozialerhebung (Version 1-0-0).

■ **Daten- und Methodenbericht:**

Daniel, A., Sarcletti, A. & Vietgen, S. (2017). 20. Sozialerhebung. Daten- und Methodenbericht zur Studierendenbefragung 2012. Version 1.0.0. Hannover: FDZ-DZHW.

Zusätzlich ist im Text mit folgender Formulierung auf die verwendeten Daten zu verweisen:

*Diese Arbeit nutzt die Daten der 20. Sozialerhebung (2012). Die Daten wurden vom Forschungsdatenzentrum des Deutschen Zentrums für Hochschul- und Wissenschaftsforschung (FDZ-DZHW) unter der doi: 10.21249/DZHW:ssy20:1.0.0 veröffentlicht.<sup>12</sup>*

---

<sup>12</sup> Bei englischsprachigen Publikationen: *This scientific work uses data of the 20th Social Survey (2012), conducted by the German Center for Higher Education Research and Science Studies (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung; DZHW). The data were published by the research data center of the DZHW, doi: 10.21249/DZHW:ssy20:1.0.0.*

## 1 Inhalt und Anlage der Studie

**[Studienreihe]** Seit 1951<sup>13</sup> führt das Deutsche Studentenwerk die Sozialerhebung durch, in der – meist in dreijährigem Abstand – Studierende standardisiert zu ihrem Studium befragt werden. Die Sozialerhebung ist damit die am längsten in Deutschland bestehende sozialwissenschaftliche Studie aus dem Bereich der Studierendenforschung. Der Fokus liegt stets auf der wirtschaftlichen und sozialen Lage der Studierenden. Die Studierenden der jeweiligen Stichprobe werden jeweils nur einmal befragt, das heißt es handelt sich um eine Querschnittsbefragung.

Die Grundgesamtheit besteht seit der 13. Sozialerhebung (1991) aus den Studierenden, die im Sommersemester des jeweiligen Erhebungsjahres an einer staatlichen oder staatlich anerkannten Hochschule in der Bundesrepublik Deutschland eingeschrieben sind. Studierende der Verwaltungsfachhochschulen, der Hochschulen des Fernstudiums und der Universitäten der Bundeswehr gehören nicht zur Grundgesamtheit.

Die ersten sechs Sozialerhebungen (1951 bis 1963) waren Vollerhebungen (ohne beurlaubte und ausländische Studierende). Seit der 7. Sozialerhebung wird jeweils nur eine Stichprobe gezogen. Seit der 10. Sozialerhebung (1982) werden die Erhebungen durch das DZHW durchgeführt. Seitdem werden auch Bildungsausländer(innen), das heißt ausländische Studierende ohne in Deutschland erworbene Hochschulzugangsberechtigung, mit in die Grundgesamtheit einbezogen (Middendorff, 2016, S. 5). Zuvor war die Sozialerhebung auf Studierende mit deutscher Staatsangehörigkeit und Bildungsinländer(innen)<sup>14</sup> beschränkt.<sup>15</sup> Seit der 17. Sozialerhebung (2003) wird die Stichprobe der Bildungsausländer(innen) als gesonderte Stichprobe gezogen und die Bildungsausländer(innen) werden seitdem mit einem eigenen zweisprachigen (deutsch und englisch) Fragebogen befragt (Middendorff, 2016, S. 5). In der 20. Sozialerhebung wurde das Spektrum der angebotenen Sprachen bei der Befragung ausländischer Studierender erweitert, da deren Befragung im Gegensatz zur 17. bis 19. Sozialerhebung nicht mehr mittels Papierfragebogen, sondern online erfolgte. Bei den zu Beginn der Online-Befragung auswählbaren Sprachen handelt es sich um Deutsch, Englisch, Französisch, Spanisch, Russisch, Chinesisch und Arabisch (Middendorff, 2016, S. 5).

**[Stichprobe der 20. Sozialerhebung]** Für die hier beschriebene 20. Sozialerhebung wurde eine Stichprobe aus der Gruppe der Studierenden mit deutscher Staatsangehörigkeit und Bildungsinländer(innen) sowie eine gesonderte Stichprobe aus der Gruppe der Bildungsausländer(innen) gezogen. Aufgrund der Unterschiedlichkeit von Erhebungsinstrument und untersuchter Studierendengruppe werden im hier dokumentierten SUF (bzw. CUF) nur die Daten der erstgenannten Stichprobe berücksichtigt.

Bis einschließlich zur 17. Sozialerhebung (2003) wurde die Sozialerhebung ausschließlich mit einem Papierfragebogen durchgeführt. In der 18. und 19. sowie der hier vorliegenden

---

<sup>13</sup> Für einen detaillierten Überblick über die Geschichte der Sozialerhebung siehe Middendorff (2016).

<sup>14</sup> Bildungsinländer(innen) sind Studierende mit ausländischer Staatsangehörigkeit, die ihre Hochschulzugangsberechtigung in Deutschland erworben haben.

<sup>15</sup> In der 6. Sozialerhebung wurde zum ersten Mal versucht, ausländische Studierende mit in die Studie einzubeziehen. Aufgrund mangelnder Datenqualität wurden die Daten jedoch nicht ausgewertet und die Befragung ausländischer Studierender bis einschließlich der 9. Sozialerhebung (1979) wieder eingestellt Middendorff (2016, S. 5).

20. Sozialerhebung wurde im Rahmen eines Methodentests<sup>16</sup> ein Teil der Stichprobe<sup>17</sup> dazu genutzt, die Studierenden online zu befragen (Middendorff, 2016, S. 12). In der 20. Sozialerhebung wurden im Rahmen eines Methodentests zwei Drittel der Zielpersonen gebeten, den Papierfragebogen auszufüllen, ein Sechstel wurde gebeten, die Befragung online durchzuführen und ein weiteres Sechstel hatte die Wahl zwischen beiden Befragungsmodi (Middendorff et al. 2013: 44). Die Daten aller Personen, die im Rahmen des Methodentests der 20. Sozialerhebung online an der Befragung teilgenommen haben, waren keine Grundlage für die Berichterstattung zum Projekt und sind nicht Bestandteil des hier dokumentierten SUF bzw. CUF. Somit sind in das hier dokumentierte SUF und CUF der 20. Sozialerhebung nur die Daten der Bildungsinländer(innen) und Studierenden mit deutscher Staatsangehörigkeit einbezogen, die mittels des Papierfragebogens befragt wurden. Von den 15.128 Befragten im Datensatz entstammen 12.859 (85,0 Prozent) der Substichprobe, die ausschließlich den Papierfragebogen erhalten hat und 2.269 (15,0 Prozent) Personen der Substichprobe, die die Wahl hatte, den Papierfragebogen auszufüllen oder online an der Befragung teilzunehmen (und die sich hierbei für die schriftliche Papierfragebogen-Option entschieden hatten).

**[Analysepotential]** Zum Kernbestand der Sozialerhebung gehören Fragen zum Hochschulzugang, zu Strukturmerkmalen des Studiums und Studienverlaufs, zur sozialen und wirtschaftlichen Lage (Studienfinanzierung, Lebenshaltungskosten, Erwerbstätigkeit, Wohnsituation), zu Themen des Tätigkeitsfeldes der Studentenwerke (*Student Services*) sowie zu soziodemographischen Merkmalen, die in Ergänzung zur amtlichen Statistik erhoben werden (Familienstand, Elternschaft, soziale Herkunft, Migrationshintergrund). Zu den Themenbereichen des Kernbestandes sind Zeitreihenanalysen möglich. Die Schwerpunktsetzung auf Themen aus dem Bereich der wirtschaftlichen und sozialen Lage der Studierenden geschieht aufgrund der Annahme, dass neben den Bedingungen des Hochschulzugangs und den Studienbedingungen auch die sozialen und wirtschaftlichen Rahmenbedingungen von großer Bedeutung für den Studienverlauf und -erfolg sind (Apolinarski et al., 2014, S. 1). Seit der 10. Sozialerhebung werden in jeder Befragung jeweils ein bis zwei ergänzende Schwerpunkte gesetzt. Hierbei handelt es sich um Themen wie z. B. Studium mit Kind, Studium mit Behinderung oder chronischer Krankheit. Die Daten der Sozialerhebung sind auch für Analysen zum Zeitaufwand für Studium und Erwerbstätigkeit geeignet. Auswertungen zum BAföG-Bezug, zu studentischer Erwerbstätigkeit und zu studienbezogenen Auslandsaufenthalten sind ebenfalls möglich. In eingeschränktem Maße sind mit den Daten auch Analysen zu Studierenden mit Kind(ern) sowie zu Studierenden mit Migrationshintergrund möglich.

**[Einordnung ins Forschungsfeld]** Bei der Sozialerhebung handelt es sich aus mehreren Gründen um eine für die Studierendenforschung einmalige Studie. Zunächst ist hier die seit 1951 bestehende Zeitreihe zu nennen, so dass mit den Daten der Sozialerhebung Fragestellungen im Bereich des Kernbestands der Studie über einen sehr langen Zeitraum hinweg untersucht werden können. Des Weiteren handelt es sich um die einzige bundesweite Studierendenstichprobe, die geeignet ist, differenzierte Aussagen über die Studierenden aller

<sup>16</sup> Da die Methodentests verschiedene Vorteile der Umstellung auf eine Online-Befragung aufgezeigt haben (vgl. Middendorff (2016, S. 12)), wird die Sozialerhebung seit 2016 (21. Sozialerhebung) ausschließlich als Online-Befragung durchgeführt.

<sup>17</sup> In die Stichprobe der Studierenden mit deutscher Staatsangehörigkeit und Bildungsinländer(innen) wird jeder 27. im Sommersemester 2012 immatrikulierte Studierende einbezogen. Für die Stichprobe der Bildungsausländer(innen) wird jeder 18. im Sommersemester 2012 immatrikulierte Studierende zufällig ausgewählt.

Semester, verschiedenster Abschlussarten, unterschiedlichster Hochschulzugangsvoraussetzungen und – mit Ausnahme der Verwaltungsfachhochschulen, der Hochschulen des Fernstudiums und der Universitäten der Bundeswehr – aller Hochschularten zu treffen.

**[Spezifika der 20. Sozialerhebung]** Im Zuge der Umsetzung der Studienstrukturreform im Rahmen des Bologna-Prozesses ist der Anteil der Studierenden in den „neuen“ Studiengängen (Bachelor, Master) von 13 Prozent in der 18. Sozialerhebung (2006) auf 47 Prozent in der 19. Sozialerhebung (2009) gestiegen (Isserstedt, Middendorff, Kandulla, Borchert & Leszczensky, 2010, S. 6). Auch in der 20. Sozialerhebung ist der Anteil (bezogen auf die Studierenden im Erststudium) mit 74 Prozent weiter gestiegen (Middendorff et al. 2013: 6).

Während der Großteil der Studierenden 2009 (19. Sozialerhebung) an einer Hochschule in den sechs Bundesländern eingeschrieben war, an der allgemeine Studiengebühren oder -beiträge erhoben wurden, wurden im Sommersemester 2012 nur noch in Bayern und Niedersachsen Studiengebühren/-beiträge erhoben (Middendorff et al. 2012: 3). Dementsprechend wurden auch die Fragen zur Höhe und Finanzierung der Studiengebühren im Rahmen der 20. Sozialerhebung nicht mehr gestellt, so dass mit den Daten der 20. Sozialerhebung Analysen zur Finanzierung oder den Auswirkungen von Studiengebühren/ -beiträgen nicht möglich sind.

Des Weiteren wurde im Zeitraum 2009 bis 2012 die Verkürzung der Schulzeit bis zum Abitur („G8“) in mehreren Bundesländern weiter umgesetzt und die Wehrpflicht wurde 2011 bundesweit ausgesetzt (Middendorff et al. 2013: 3), was dazu führt, dass das Alter, in dem ein Studium frühestens aufgenommen werden kann, gesunken ist.

## 2 Erhebungsinstrument

Für die hier dokumentierte 20. Sozialerhebung wurde ein standardisierter Papierfragebogen in deutscher Sprache als Erhebungsinstrument eingesetzt.<sup>18</sup> Kapitel 2.1 stellt die zentralen Inhalte des Erhebungsinstrumentes vor. Kapitel 2.2 beschreibt den zur Prüfung und Verbesserung des Fragebogens durchgeführten Pretest.

### 2.1 Inhalte der Erhebungsinstrumente

**[Charakteristika der Studienreihe]** Im Fokus der 20. Sozialerhebung steht, wie bei den übrigen Erhebungen der Studienreihe, die soziale und wirtschaftliche Situation der Studierenden. Die Kernthemen sind dabei Strukturmerkmale des Studiums, die finanzielle Situation der Studierenden, ihre soziale Herkunft, sowie demographische Angaben. Diese Kernthemen werden in allen Erhebungen abgefragt. Daneben werden meist ein oder mehrere Schwerpunktthemen (z. B. Computer- und Internetnutzung, Beratungs- und Informationsbedarf) erfasst.

Die Fragen bezüglich der Strukturmerkmale des Studiums erfassen Informationen zum derzeitigen Studium (z. B. Studienfach, angestrebter Abschluss, Hochschul- und Fachsemester), zum Studienverlauf (z. B. bereits erworbene Hochschulabschlüsse, Studienunterbrechungen, Fach- und Hochschulwechsel; in der 20. Sozialerhebung: Fragen 1 bis 11) sowie Angaben zum zeitlichen Aufwand für Studium und Erwerbstätigkeit (Frage 13). Zudem werden studienbezogene Auslandserfahrungen und deren Finanzierung, Hinderungsgründe bezüglich eines (Teil-)Studiums im Ausland (Fragen 49 bis 51) sowie Sprachkenntnisse (Frage 52) erfasst. Diese Angaben werden durch persönliche Einschätzungen zum Stellenwert des Studiums und zur subjektiv empfundenen zeitlichen Belastung ergänzt (Fragen 12 und 14).

Im Fokus jeder Sozialerhebung steht die finanzielle Situation der Studierenden. Ausgaben für Lebenshaltung werden betragsgenau erfragt (Frage 20). Ebenfalls betragsgenau erfasst werden die Einnahmen aus eigener Erwerbstätigkeit, familiärer/partnerschaftlicher Unterstützung, Stipendien und weiteren Quellen (Frage 19). Detaillierte Angaben zur Nutzung der Ausbildungsförderung durch das Bundesausbildungsförderungsgesetz (BAföG) werden zusätzlich erhoben (Fragen 22 bis 25). Angaben zu Erwerbstätigkeiten im aktuellen Semester und in der vorlesungsfreien Zeit vor dem Start des aktuellen Semesters (Frage 26, Frage 27.1 und Frage 27.2) werden durch Items zur subjektiven Einschätzung der finanziellen Situation (Frage 21) und zu Motiven der eigenen Erwerbstätigkeit (Frage 27.3) erweitert. Darüber hinaus werden verschiedene soziodemographische (Fragen 28 und 29 sowie Fragen 33 bis 40) und bildungsbiographische Merkmale (Fragen 15 bis 18) sowie der elterliche Hintergrund (Fragen 42 bis 48) erhoben. Hinzu kommen Fragen zur Nutzung von Mensen und Cafeterien (Fragen 30 bis 32).

**[Spezifika der 20. Sozialerhebung]** Aufgrund der weit fortgeschrittenen Umstellung der Studienstruktur wurden in der 20. Sozialerhebung (wie auch schon in der 19. Sozialerhebung) Fragen zum Masterstudium (Fragen 6.1 bis 6.4) ergänzt.

Ein Schwerpunktthema, das in den letzten beiden Sozialerhebungen (2006 und 2009) nicht behandelt wurde, ist die Computer- und Internetnutzung der Studierenden. Hier werden

<sup>18</sup> Der Fragebogen kann von der Website des FDZ heruntergeladen werden. Ebenso steht dort ein Filterführungsdiagramm des Fragebogens zur Verfügung.

die Sicherheit im Umgang mit verschiedenen Computer- und Internetanwendungen (Frage 53) und Gründe für die Nutzung von Computer und Internet in Freizeit und Studium (Frage 54) erhoben.

Zusätzlich werden Fragen zu gesundheitlichen Beeinträchtigungen gestellt, die lange Zeit zu den Kernthemen zählten, aber zuletzt in der 17. und 19. Sozialerhebung nicht erhoben wurden (Fragen 41.1 bis 41.6). Dabei geht es neben der Art der Beeinträchtigung auch um eventuelle Ausgaben für Studien- und Kommunikationsassistenzen. Allerdings gehören Daten zur Gesundheit zu den „besonderen Arten personenbezogener Daten“<sup>19</sup> und sind daher besonders schützenswert (vgl. §3 Abs. 9 BDSG, Art. 8 Abs. 1 und 2a EG-DSRL), so dass eine Bereitstellung dieser Daten aus rechtlichen Gründen nicht möglich ist (siehe auch Kapitel 8).

Ergänzend zur Staatsangehörigkeit der Studierenden (seit der 14. Sozialerhebung) und der Staatsangehörigkeit der Eltern der Studierenden, die seit der 19. Sozialerhebung vorliegt, wurde in der 20. Sozialerhebung erfragt, ob die Studierenden (Frage 40) und deren Eltern (Frage 48) in Deutschland geboren wurden. Dadurch ist in der 20. Sozialerhebung eine genauere Erfassung des Migrationshintergrundes als bei früheren Sozialerhebungen möglich. Im Gegensatz zur 17. bis 19. Sozialerhebung sind in der 20. Sozialerhebung keine Informationen zum Beratungs- und Informationsbedarf der Studierenden abgefragt worden.

## 2.2 Pretest

**[Ziel und Verfahren]** Das Erhebungsinstrument wurde im Vorfeld der Erhebung durch einen kognitiven Pretest geprüft. Dabei sollte erstens für die bereits in Erhebungsinstrumenten vorheriger Sozialerhebungen eingesetzten Fragen und Antwortvorgaben geprüft werden, ob sie von den Studierenden 2012 gleich perzipiert werden würden wie von den Studierenden vorangegangener Sozialerhebungen. Zweitens sollte für die neu eingesetzten oder modifizierten Fragen die Verständlichkeit und Beantwortbarkeit getestet sowie ermittelt werden, wie Studierende bei der Beantwortung ausgewählter Fragen (z. B. Zeitbudget, Einnahmen, Ausgaben) vorgehen. Die Aufnahme der neuen Fragen zog außerdem Veränderungen am Aufbau und Layout des Fragebogens sowie der Befragungsdauer nach sich, die evaluiert werden sollten. Zur Prüfung dieser verschiedenen Aspekte kam ein sogenanntes „Pretestverfahren im Feld“ zum Einsatz. Dieses Verfahren hat zum Ziel, dass die am Pretest teilnehmenden Personen „unter möglichst ähnlichen Bedingungen untersucht [werden], wie sie später für die eigentliche Erhebung vorgesehen sind“ (Häder, 2015, S. 396).

**[Probanden]** Über öffentlich zugängliche Aushänge an der örtlichen Universität (Leibniz Universität Hannover) wurden insgesamt 31 Studierende mit in Deutschland erworbener Hochschulzugangsberechtigung zur Teilnahme an insgesamt drei Pretest-Sitzungen rekrutiert. Von diesen Personen hatte eine zwei Staatsbürgerschaften. Drei Studierende hatten die Staatsbürgerschaft gewechselt. Bei den übrigen Studierenden handelt es sich um Personen ausschließlich mit deutscher Staatsbürgerschaft, die die Staatsbürgerschaft nicht gewechselt haben.

---

<sup>19</sup> Neben Angaben zur Gesundheit gehören auch Angaben über die rassistische und ethnische Herkunft, politische Meinungen, religiöse oder philosophische Überzeugungen, Gewerkschaftszugehörigkeit oder Sexualleben zu den besonderen Arten personenbezogener Daten (§3 Abs. 9 BDSG).



**[Durchführung]** Der Pretest fand in drei Sitzungen im Januar 2012 und damit etwa vier Monate vor dem Feldstart statt. Dabei wurden die Testpersonen gebeten, den für die Befragung vorgesehenen standardisierten Fragebogen zu beantworten. Im Anschluss daran führte ein(e) Mitarbeiter(in) der Sozialerhebung mit den Studierenden in einer Gruppendiskussion (Fokusgruppen mit jeweils etwa zehn Teilnehmer(inne)n pro Sitzung) anhand eines Leitfadens ein Auswertungsgespräch durch. Der Fokus dieses Gesprächs lag dabei auf den neu eingesetzten oder modifizierten Fragen, wobei die Studierenden die Möglichkeit hatten, sich zu jeder Frage zu äußern. Zusätzlich erfolgten Nachfragen zur Ausfülldauer, zu Inhalt und Länge des Fragebogens, zu Aufbau und Layout, zur Verständlichkeit der Fragen und Ausfüllanweisungen sowie zur Vollständigkeit der Antwortmöglichkeiten. Auf Grundlage der Pretestergebnisse wurden einige Fragen leicht umformuliert, erweitert oder durch zusätzliche Ausfüllhinweise ergänzt. Der grundsätzliche Aufbau, der Umfang und das Layout der Fragebögen wurden unverändert beibehalten.

### 3 Grundgesamtheit und Stichprobenverfahren

**[Grundgesamtheit]** Die Grundgesamtheit der 20. Sozialerhebung umfasst Studierende mit deutscher Staatsangehörigkeit und Bildungsinländer(innen)<sup>20</sup>, die im Sommersemester 2012 an einer staatlich anerkannten Hochschule in der Bundesrepublik Deutschland eingeschrieben waren. Studierende an Verwaltungsfachhochschulen, Hochschulen des Fernstudiums sowie Studierende an den Universitäten der Bundeswehr waren dabei ausgenommen, da sich ihre Studiensituation grundsätzlich von der Studiensituation anderer Studierender unterscheidet.<sup>21</sup> Es ergibt sich gemäß der Studierendenstatistik des Statistischen Bundesamtes für das Wintersemester 2011/2012 eine Größe der Grundgesamtheit von 2.035.217 Studierenden.

Für kleine Hochschulen steht der entstehende Verwaltungsaufwand in einem sehr schlechten Verhältnis zur Zahl der auswertbaren Fragebögen, die man durch die Teilnahme der Hochschule erzielen könnte, so dass Hochschulen mit weniger als 110 immatrikulierten Studierenden ausgeschlossen wurden. Dies trifft bei 41 Hochschulen zu. Bei diesen Hochschulen wurde folglich auf eine Teilnahmebitte verzichtet. Von den angefragten 331 Hochschulen nahmen 227 an der Studie teil (68,6 Prozent). An den teilnehmenden Hochschulen waren im Wintersemester 2010/2011 90,7 Prozent der Studierenden (Deutsche und Bildungsinländer(innen)) der oben definierten Grundgesamtheit eingeschrieben.

**[Stichprobenverfahren]** Aufgrund nicht zugänglicher Listen von Studierenden war es ausgeschlossen, eine einfache Zufallsstichprobe zu ziehen. Es wurde eine Zufallsstichprobe auf Ebene aller Hochschulen konzipiert, an denen die zur Grundgesamtheit gehörenden Studierenden immatrikuliert waren, da diese nur über die Hochschulen erreicht werden konnten. An den relevanten Hochschulen wurde aus der Gruppe von Studierenden mit deutscher Staatsangehörigkeit und Bildungsinländer(inne)n eine Zufallsstichprobe mit der Auswahlwahrscheinlichkeit von 1/27 (d. h. jede(r) 27. im Sommersemester 2012 immatrikulierte Studierende) gezogen.<sup>22</sup> An einzelnen Hochschulen wurde auf Anfrage der Studentenwerke eine größere Stichprobe (d. h. mit einer erhöhten Auswahlwahrscheinlichkeit) gezogen, um auch Auswertungen auf Ebene der Studentenwerke zu ermöglichen (disproportionale Ziehung).<sup>23</sup> Die disproportionale Ziehung wird durch eine entsprechende Designgewichtung ausgeglichen (siehe Kapitel 7).

<sup>20</sup> Bildungsinländer(innen) sind Studierende mit ausländischer Staatsangehörigkeit, die ihre Hochschulzugangsberechtigung in Deutschland erworben haben.

<sup>21</sup> Studierende an Fernhochschulen sind sehr häufig berufstätig und absolvieren das Studium nebenbei. Studierende an Verwaltungsfachhochschulen sind in der Regel Beamtenanwärter(innen) („Beamte auf Widerruf“). Studierende an Universitäten der Bundeswehr sind in der Regel Offiziersanwärter(innen).

<sup>22</sup> Zum Zweck der internen methodischen Weiterentwicklung der Studie erhielten ein Sechstel der Stichprobe statt des Papierfragebogens einen Link zur Online-Version der Befragung. Ein weiteres Sechstel der Stichprobe konnte zwischen der Online-Befragung und der Papierbefragung wählen. Alle Studierenden, die online an der Befragung teilgenommen haben, sind nicht Bestandteil des SUF und werden daher in dieser Dokumentation nicht berücksichtigt (vgl. Kapitel 1).

<sup>23</sup> Stichprobenerhöhungen wurden an folgenden 17 Hochschulen durchgeführt: Universität Hannover, Medizinische Hochschule Hannover, Tierärztliche Hochschule Hannover, Hochschule für Musik, Theater und Medien Hannover, Hochschule Hannover und Fachhochschule für die Wirtschaft Hannover (jeweils jede(r) 21.), Karlsruher Institut für Technologie, Pädagogische Hochschule Karlsruhe und Hochschule Karlsruhe (jeweils jede(r) 20.), Universität Tübingen (jede(r) 17.), Technische Universität Darmstadt (jede(r) 16.), Universität Kassel und Universität Marburg (jede(r) 15.), Universität Wuppertal (jede(r) 11.), Universität Ulm und Hochschule Ulm (jede(r) 10.), Hochschule Darmstadt (jede(r) 9.)

## 4 Durchführung der Erhebung

**[Kontaktaufnahme]** Alle Hochschulen mit Studierenden der Grundgesamtheit wurden schriftlich vom DZHW kontaktiert und um Teilnahme an der 20. Sozialerhebung gebeten. Im Anschreiben wurde auf die Relevanz der Studie und ihre Geschichte aufmerksam gemacht. Der Brief an die Hochschulen enthielt darüber hinaus ein Schreiben der Präsidentin der Hochschulrektorenkonferenz (HRK) mit der Bitte um Teilnahme an der Sozialerhebung. Zudem teilte das DZHW den Hochschulen die Kriterien mit, anhand derer sie die Zielpersonen für die 20. Sozialerhebung identifizieren sollten.<sup>24</sup> 227 der 331 angeschriebenen Hochschulen (68,6 Prozent) erklärten sich zur Teilnahme bereit.<sup>25</sup> Da die Hochschulen die Kontaktdaten ihrer Studierenden aus Datenschutzgründen nicht herausgeben dürfen, werden im Rahmen der Sozialerhebungen die Erhebungsunterlagen durch die Hochschulen versendet. Das DZHW ermittelte dafür jeweils die Anzahl der benötigten Erhebungsunterlagen<sup>26</sup> und sendete diese an die teilnahmebereiten Hochschulen, die aus ihrem Studierendenverzeichnis die Adressen der Studierenden zogen und sie jeweils zwei Mal auf Adresstiketten druckten. Ein Etikett wurde für den Umschlag mit den Erhebungsunterlagen benötigt, ein zweites für die Erinnerungspostkarte. Beides verschickten die Hochschulen jeweils in einem vorgegebenen Zeitfenster an die Zielpersonen.

**[Erhebungsunterlagen]** Die Erhebungsunterlagen bestanden pro zu befragender Person aus einem Anschreiben des Deutschen Studentenwerks als Auftraggeber der Studie, dem Fragebogen, einem Flyer mit allgemeinen Informationen zur Sozialerhebung sowie exemplarischen Ergebnissen vorheriger Erhebungen und einem an das DZHW adressierten Freiumschlag zur Rücksendung des ausgefüllten Fragebogens. Darüber hinaus erhielten alle Studierenden der gezogenen Stichprobe zeitversetzt eine Erinnerungspostkarte.

**[Feldphase]** Der Erhebungszeitraum erstreckte sich vom 29. Mai bis 20. August 2012. Aufgrund des angewendeten Kontaktverfahrens über die Hochschulen konnte das DZHW keinen direkten Einfluss auf den genauen Versandzeitpunkt der Erhebungsunterlagen nehmen. Die Erinnerungspostkarten wurden etwa zwei Wochen nach Feldstart an die Studierenden verschickt. Sie enthielten auch eine Dankesformel, denn sie wurden an alle Personen der Stichprobe verschickt – also auch diejenigen, die sich bereits an der Befragung beteiligt hatten –, da den Hochschulen unbekannt war, welche Personen bereits einen Fragebogen an das DZHW zurückgeschickt hatten.

**[Rücklaufsteigernde Maßnahmen]** Neben der Erinnerungspostkarte und dem zusammen mit dem Anschreiben verschickten Flyer wurde die 20. Sozialerhebung durch Presse- und Öffentlichkeitsarbeit seitens des Deutschen Studentenwerks und der einzelnen Studierendenwerke beworben. Es wurde mit Pressemitteilungen und Plakatwerbung auf die bevorstehende Befragung hingewiesen. Darüber hinaus stand die Projektwebsite (<http://www.sozialerhebung.de>) als Informationsquelle zur Verfügung. Sie enthielt die Frage-

<sup>24</sup> In der Regel (vgl. Kapitel 3) jeder 27. im Sommersemester immatrikulierte Studierende mit deutscher Staatsangehörigkeit oder Bildungsinländer(in).

<sup>25</sup> Kleinere, Kunst- und Musik- sowie private Hochschulen gaben überproportional keine positive Rückmeldung. Bezogen auf die Studierendenzahl sind dennoch knapp 91 Prozent an den 227 (von 331) Hochschulen, die sich zur Teilnahme bereit erklärt haben, immatrikuliert.

<sup>26</sup> Dazu wurden die Zahlen der amtlichen Statistik zu Studierenden an den Hochschulen des vorangegangenen Semesters (Wintersemester 2011/12) mit der jeweiligen Auswahlwahrscheinlichkeit kombiniert.

bögen, Fotos und Textbausteine für Pressemitteilungen und weiteres Informationsmaterial zur aktuellen Studie sowie zu vorangegangenen Erhebungen.<sup>27</sup> Zusätzlich wurde ein Facebook-Auftritt als Serviceportal eingerichtet ([www.facebook.com/Sozialerhebung](http://www.facebook.com/Sozialerhebung)).

## 5 Rücklauf

**[Rücklauf]** Die Zahl der verschickten Befragungsunterlagen liegt bei etwa 69.814<sup>28</sup>, da ungefähr diese Anzahl an Personen von den Hochschulen als Stichprobe gezogen wurde. In 1.471 Fällen gab es unzustellbare Sendungen. Wenn dem DZHW bekannt wurde, dass Studierende angeschrieben wurden, die nicht Bestandteil der Grundgesamtheit sind, wurden diese Fälle (n=123) – genauso wie die unzustellbaren Fragebögen – als stichprobenneutrale Ausfälle gewertet. Daher haben etwa 68.220 zur Stichprobe gehörende Studierende die Einladung zur Teilnahme erhalten. Circa<sup>29</sup> zwei Drittel der angeschriebenen Studierenden (ungefähr 45.844) erhielten die Einladung zur Teilnahme an der papierbasierten Befragung.<sup>30</sup> Von diesen wurden 12.859 Fragebögen an das DZHW zurückgeschickt. Somit verblieben in der Nettostichprobe 12.859 Fälle. Im Hinblick auf die Bruttostichprobe liegt die Rücklaufquote damit bei rund 28 Prozent (vgl. Tabelle 1).

**Tabelle 1: Brutto- und Nettostichprobe sowie Rücklaufquote der 20. Sozialerhebung (2012)**

Bruttostichprobe	45.844
Nettostichprobe	12.859
Rücklaufquote	28,0 %

Abbildung 2 (auf der nächsten Seite) stellt den Rücklauf der Fragebögen im Zeitverlauf dar. Wie sich zeigt, hat ein Großteil der ausgefüllten Fragebögen das DZHW während der ersten Hälfte der Feldphase erreicht, in der auch die Erinnerungspostkarte verschickt wurde. Insgesamt erstreckte sich die Rücklaufphase über elf Wochen.

**[Zusätzliche Fälle]** Etwa ein Sechstel der Befragten erhielt Befragungsunterlagen, die sowohl die Teilnahme an der schriftlichen Befragung als auch eine Online-Teilnahme ermöglichten. Nach Abzug der stichprobenneutralen Ausfälle haben etwa 11.188 Studierende die Wahl zwischen schriftlicher und Online-Teilnahme gehabt. 2.980 Personen (26,6 Prozent) haben an

<sup>27</sup> Die Materialien zur 20. Sozialerhebung sind unter <http://www.sozialerhebung.de/sozialerhebung/archiv> (Stand: 01.06.2017) zu finden.

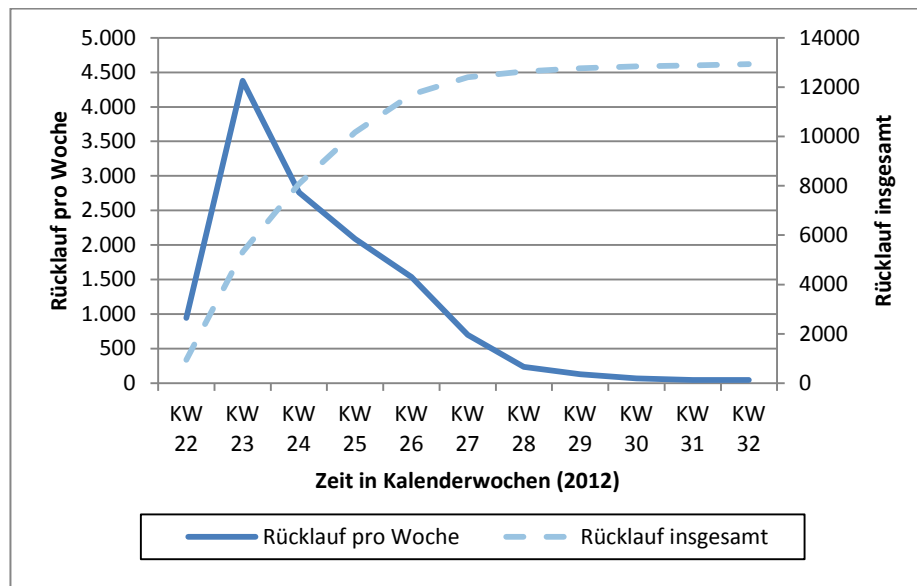
<sup>28</sup> 192 der 227 teilnehmenden Hochschulen teilten die tatsächliche Zahl der angeschriebenen Studierenden mit. Die angegebene Zahl beruht auf einer Hochrechnung der Rückmeldung dieser 192 Hochschulen auf alle 227 teilnehmenden Hochschulen.

<sup>29</sup> Die exakte Anzahl an verschickten Erhebungsunterlagen für die drei Varianten (Papierfragebogen; Online-Befragung; Wahl zwischen Online-Befragung und Papierfragebogen) ist nicht rekonstruierbar. Den Mitarbeiter(inne)n der Hochschulen war nicht bekannt, in welchen Kuverts sich welche Version der Erhebungsunterlagen befand. Teilweise blieben auch Erhebungsunterlagen nach der Etikettierung übrig. Da die Hochschulen jedoch annähernd passend mit der Anzahl der Erhebungsunterlagen ausgestattet waren, die sie auf Grundlage ihrer Stichprobe ermittelt hatten, ist davon auszugehen, dass die Unterlagen etwa in dem Verhältnis verschickt wurden, dass zwei Drittel eine Einladung zur schriftlichen Teilnahme, ein Sechstel zur Online-Teilnahme und ein weiteres Sechstel eine Einladung mit beiden Optionen erhielten.

<sup>30</sup> Da eine genaue Differenzierung zwischen einer (unbereinigten) Brutto-Ausgangsstichprobe und einer bereinigten Bruttostichprobe bezogen auf die drei Erhebungsmodi nicht möglich ist, wird sich im Folgenden auf die Bezeichnung „Bruttostichprobe“ beschränkt.

der Befragung teilgenommen, wobei 711 (6,4 Prozent) online teilgenommen haben und 2.269 (20,3 Prozent) den Fragebogen ausgefüllt haben. Die Daten der Personen aus dieser Teilstichprobe, die den Papierfragebogen ausgefüllt haben, sind Bestandteil des Datensatzes, die Daten derer, die an der Online-Befragung teilgenommen haben, hingegen nicht. Insgesamt sind im Datensatz somit  $(12.859 + 2.269 =)$  15.128 Fälle enthalten.

**Abbildung 2: Rücklauf der 20. Sozialerhebung (2012) im Zeitverlauf**



## 6 Datenaufbereitung

Im Folgenden werden die verschiedenen Schritte der Datenaufbereitung beschrieben. Die in den Kapiteln 6.1 bis 6.3 beschriebenen Tätigkeiten wurden bereits durch das Primärforschungsprojekt durchgeführt. Die Generierung von Variablen (Kapitel 6.4) wurde sowohl durch das Primärforschungsprojekt als auch im Rahmen der Datenedition durch das FDZ-DZHW vorgenommen. Die in den Kapiteln 6.5 bis 6.7 dargestellten Tätigkeiten wurden durch das FDZ durchgeführt, teilweise aufbauend auf Vorarbeiten des Primärforschungsprojektes. Die im Rahmen der Datenedition vorgenommenen Aufbereitungsprozesse der Gewichtung und Anonymisierung werden in den beiden folgenden Kapiteln 7 und 8 gesondert erläutert.

### 6.1 Datenübertragung

Zur weiteren Verarbeitung wurden die Angaben der Befragten aus den Papierfragebögen auf Basis eines Codeplans in ein computerlesbares Format übertragen. Zuvor wurden auf den Papierfragebögen numerische Codierungen für die offenen Angaben vermerkt (vgl. Kapitel 6.2) sowie manuelle Vorkorrekturen zur Erleichterung der Datenübertragung vorgenommen (vgl. Kapitel 6.3).

**[Erstellung eines Codeplans]** Auf Basis des Fragebogens der Befragung wurde ein Codeplan erstellt. Dabei wurde vermerkt, welcher Frage bzw. Teilfrage eine Variable zugeordnet ist, welchen Namen diese Variable trägt und welche numerischen Codierungen für die standardi-

sierten Antworten der Befragten verwendet werden. Um die Erfassungsreihenfolge festzulegen, wurden die Variablen zusätzlich nummeriert.<sup>31</sup>

**[Datenerfassung]** Für die Datenübertragung wurden der Codeplan, weitere Anweisungen zur Datenerfassung sowie die vorbereiteten Papierfragebögen an einen externen Dienstleister übergeben. Die Erfassung der Angaben erfolgte dort manuell durch Schreibkräfte. Die Qualität der Datenerfassung wurde stichprobenartig geprüft. Im Rahmen der Datenprüfung und -bereinigung (s. Abschnitt 6.3) fielen Erfassungsfehler, die durch Überspringen von Variablen entstanden, meist durch Wertebereichsverletzungen in nachgelagerten Variablen auf und wurden entsprechend korrigiert.

## 6.2 Codierung offener Angaben

Vor der Datenerfassung erfolgte eine Codierung eines Teils der (halb-)offenen Angaben. Dabei wurden diesen anhand einer Codierliste numerische Codierungen zugeordnet. Je nach Variable wurden unterschiedliche Codierlisten verwendet. Es handelt sich um Klassifikationsschlüssel der amtlichen Statistik (z. B. Schlüsselverzeichnis der Studenten- und Prüfungsstatistik oder amtliche Codes für Staaten) oder um bereits in vorherigen Sozialerhebungen eingesetzte Schlüssel. Für wenige Variablen wurden neue Codierlisten auf Basis der in den Daten der 20. Sozialerhebung vorkommenden Nennungen entwickelt. Für einige halboffene Fragen wurden keine neuen Variablen mit numerischen Codierungen erstellt, sondern die Nennungen nur – sofern möglich – den vorhandenen (geschlossenen) Antwortkategorien zugeordnet.<sup>32</sup> Einzelne offene Fragen wurden nicht vercodet. Dies war der Fall, wenn

- sie für die Auswertungen im Rahmen der 20. Sozialerhebung nicht benötigt wurden.
- lediglich relevant war, ob überhaupt eine Nennung erfolgte.
- die Fragen durch geschlossen erfasste Fragen annähernd ersetzt werden konnten.<sup>33</sup>
- eine Vercodung aufgrund mangelnder personeller und zeitlicher Ressourcen nicht möglich war.

Die durch das Primärforschungsprojekt vorgenommenen Codierungsentscheidungen wurden unverändert beibehalten.

In

<sup>31</sup> Die Daten wurden in einem einfachen spaltenorientierten Textformat ohne eine die Variablenamen enthaltende Kopfzeile erfasst. Der Codeplan musste daher festlegen, in welcher Reihenfolge die Daten zu erfassen sind, damit die zu einer Variable zugehörigen Daten in der richtigen Spalte eingetragen werden konnten.

<sup>32</sup> Ein Beispiel betrifft Frage 27.2: Wurde hier als sonstige Tätigkeit eine Praktikumsvergütung genannt, dann wurden die Variablen job03m\_c (sonstige Tätigkeit genannt) und job04m\_c (Nettostundenlohn bei der sonstigen Tätigkeit) auf „nicht genannt“ (job03m\_c) bzw. „trifft nicht zu“ (job04m\_c) zurückgesetzt, die Variable job03g\_c (Tätigkeit SoSe 2012: vergütetes Praktikum) auf „genannt“ gesetzt und bei Variable job04g\_c (Nettostundenlohn: vergütetes Praktikum) der bei job04m\_c (Nettostundenlohn bei der sonstigen Tätigkeit) genannte Wert eingetragen.

<sup>33</sup> Die offenen Angaben zu folgenden Fragen wurden elektronisch erfasst. Es wurde jedoch auf eine Vercodung verzichtet und nur in der Variable dokumentiert, ob eine offene Angabe erfolgt war (genannt) oder nicht (nicht genannt): Frage 11.2 (persönliche Gründe, die für die Wahl der gegenwärtigen Hochschule eine Rolle spielten), Frage 15 (andere Studienberechtigung), Frage 19.1 (andere Finanzierungsquelle), Frage 24 (anderer Grund, warum bisher kein BAföG-Antrag gestellt wurde), Frage 27.2 (andere Tätigkeit gegen Bezahlung), Frage 51 (Hinderungsgrund für Auslandsstudium: sonstiger Grund) sowie Frage 45 (Berufe der Eltern). Alternativ zu den Berufen der Eltern können der Erwerbsstatus (Frage 42), der schulische (Frage 43) und berufliche (Frage 44) Abschluss sowie die berufliche Position (Frage 46) der Eltern für die Bestimmung der sozialen Herkunft verwendet werden.

Tabelle 2 sind die codierten Merkmale sowie die jeweils verwendete Codierliste dargestellt. Der Datensatz beinhaltet ausschließlich die codierten numerischen Variablen, die offenen Nennungen selbst sind nicht im Datensatz enthalten. Die Ausprägungen der einzelnen Variablen sind im Datensatzreport dokumentiert.

**Tabelle 2: Vercodete Merkmale und verwendete Codierschemata in der 20. Sozialerhebung (2012)**

Merkmalsname	Codierschema	Codierschema-ID
Studienfach	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2011/2012 und SoSe 2012)	cl-destatis-studienfach-2012
Studienabschluss	Projekteigene Codierung	cl-dzhw-14
Hochschule	Destatis Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2011/2012 und SoSe 2012)	cl-destatis-hochschule-2012
Bundesland	Destatis Bundeslandschlüssel (entsprechend der ersten beiden Ziffern des Amtlichen Gemeindecodierungsschlüssels (AGS))	cl-destatis-bundesland-1990 cl-dzhw-18
Ausland / Staatsangehörigkeit	Destatis Staatsangehörigkeits- und Gebietsschlüssel (Stand: 01.01.2012) (mit Anpassungen)	cl-dzhw-15
Sonstige offene Abfragen	Zuordnung zu vorgegebenen Kategorien oder projekteigene Codierung	

### 6.3 Datenprüfung und Datenbereinigung

**[Manuelle Vorkorrektur]** Bereits vor der Übertragung der Daten wurden auf den Papierfragebögen eine manuelle Prüfung und gegebenenfalls eine Anpassung von Angaben der Befragten durchgeführt.<sup>34</sup> Dies sollte zum einen die Erfassbarkeit der Daten erleichtern. Dafür wurde die Form, aber nicht der Inhalt der gemachten Angaben, verändert. Beispielsweise wurden schwer lesbare Angaben oder Streichungen der Befragten verdeutlicht.

Zum anderen zielte die manuelle Prüfung darauf ab, schon vor der softwaregestützten Korrektur (siehe unten) erste Fehler oder Inkonsistenzen in den Angaben der Befragten zu bereinigen. Dafür wurden vorab verschiedene Konsistenzregeln aufgestellt, die auf den Papierfragebögen zu überprüfen waren. Bei festgestellten Inkonsistenzen wurden die betroffenen Angaben möglichst plausibilisiert und bei fehlenden Angaben – sofern möglich – aus anderen Nennungen im Fragebogen rekonstruiert. Zudem erfolgten inhaltliche Konsistenzprüfungen. Beispielsweise fand ein interner Abgleich der Tätigkeiten der Eltern (Frage 42) statt: Gab ein(e) Befragte(r) an, dass die Mutter oder der Vater sowohl „Rentner(in)/Pensionär(in)“ als auch „teilzeiterwerbstätig“ ist, so wurde lediglich die Angabe „Rentner(in)/Pensionär(in)“ erfasst. Festgestellte Inkonsistenzen wurden – falls möglich – durch den Abgleich mit anderen Nennungen<sup>35</sup> im Fragebogen aufgelöst oder andernfalls ein entsprechender Missingcode (vgl. Kapitel 6.7) vergeben.

**[Softwaregestützte Korrektur]** Im Anschluss an die Datenübertragung erfolgte eine umfassende Prüfung und Korrektur der Daten mit Hilfe einer DZHW-eigenen Software. Dabei sollten zum einen Fehler bei der vorherigen manuellen Vorkorrektur und Datenübertragung,

<sup>34</sup> Die Zahl der vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebögen dokumentiert und ist daher nicht mehr systematisch rekonstruierbar.

<sup>35</sup> Ein Beispiel hierfür ist Frage 2 (angestrebter Abschluss im derzeitigen Studiengang): Hier wurden die Befragten gebeten, nur den Abschluss anzugeben, der zunächst erworben werden soll. Wurde entgegen dieser Vorgabe von Befragten mehr als eine Angabe gemacht, so wurde nur der Abschluss codiert, der üblicherweise zuerst erworben wird. Wenn z. B. „Staatsexamen (ohne Lehramt)“ und „Promotion“ angekreuzt wurden, so wurde nur die Schlüsselziffer für „Staatsexamen (ohne Lehramt)“ erfasst.



zum anderen weitere inkonsistente Angaben der Befragten, die bei der Vorkorrektur nicht geprüft werden konnten, identifiziert werden.

Zu diesem Zweck wurden die erfassten Fragebogen-Daten in eine Datenbank eingelesen. Anschließend wurden anhand formaler Regeln gültige Wertebereiche und Antwortkombinationen definiert und geprüft. Folgende Typen von Prüfungen wurden vorgenommen:

- *Prüfung von Wertebereichen:* Es wurde geprüft, ob die erfasste Ausprägung einer Variablen in dem für diese Variable definierten Wertebereich lag.
- *Prüfung der Einhaltung der Filterführung:* Es wurde auf der einen Seite geprüft, ob aufgrund der im Fragebogen vorgesehenen Filterführung für die jeweilige befragte Person Angaben zu erwarten gewesen wären (Vollständigkeitsprüfung), und auf der anderen Seite, ob für die jeweilige Person keine Angaben hätten erfolgen dürfen (Filterverstöße).
- *Konsistenzprüfung:* Es wurde die Konsistenz der Angaben innerhalb des Fragebogens geprüft. Neben verschiedenen Merkmalskombinationen, die bereits bei der manuellen Vorkorrektur überprüft wurden, erfolgte hier eine noch detailliertere Definition von Merkmalskombinationen, die erfüllt sein mussten.<sup>36</sup>

Insgesamt wurden mehrere hundert Konsistenzregeln getestet. Bei fehlenden, fehlerhaften oder unplausiblen Werten wurde zunächst mit Hilfe des Papierfragebogens geprüft, ob der entsprechende Wert falsch (bzw. nicht) übertragen worden war. Ansonsten wurde versucht, den korrekten Wert anhand anderer Angaben im Fragebogen zu erschließen. Im Zweifelsfall wurde ein spezifischer Missingcode vergeben (vgl. Kapitel 6.7). Fehlerkorrekturen wurden dokumentiert<sup>37</sup> und von mindestens einer weiteren Person geprüft.

**[Löschung von Fällen]** Fälle, die nicht zur Grundgesamtheit gehörten, wurden aus dem Datensatz gelöscht (siehe Kapitel 5).<sup>38</sup> Dies geschah nicht immer bereits im Rahmen der manuellen Vorkorrektur, so dass im Nachhinein (nach der Datenerfassung) noch 44 Befragte identifiziert wurden, die weder eine deutsche Staatsangehörigkeit hatten noch ihre Hochschulzugangsberechtigung in Deutschland erworben haben. Des Weiteren wurden im Nachhinein 24 Fernstudierende identifiziert. Insgesamt wurden 123 nicht zur Grundgesamtheit gehörende Fälle aus dem Datensatz entfernt, davon 55 im Rahmen der manuellen Vorkorrektur und 68 nach der Datenerfassung im Rahmen der Plausibilisierung.

<sup>36</sup> Besonders detailliert wurden die Einnahmen (Frage 19.1) und Ausgaben (Frage 20) sowie das Zeitbudget (Frage 13) geprüft.

<sup>37</sup> Die Dokumentation der Fehlerkorrekturen erfolgte handschriftlich auf den Papierfragebögen und ist daher nicht systematisch rekonstruierbar.

<sup>38</sup> Dies traf dann zu, wenn eine Person die Unterlagen erhalten hat, die Bildungsausländer(in) ist, exmatrikuliert oder beurlaubt ist oder zur Gruppe der Fernstudierenden gehört.

## 6.4 Generierung von Variablen

Neben den Variablen, die die codierten Antworten der Befragten enthalten, beinhaltet der Datensatz der 20. Sozialerhebung auch generierte Variablen. Dabei handelt es sich zum einen um Variablen mit numerischen Codierungen von ursprünglich offenen Nennungen (vgl. Kapitel 6.2). Zum anderen wurden Variablen aus Datenschutzgründen verändert und es wurden im Forschungsfeld häufiger benötigte Variablen aus den Werten einer oder mehrerer Quellvariablen generiert (z. B. Aggregation der Studienfächer zu Studienbereichen und Fächergruppen oder Ableitung von Hochschulart und Hochschulort aus den Hochschulvariablen). Ein großer Teil der generierten Variablen wurde bereits durch das Primärforschungsprojekt erstellt. Der Variablenname einer generierten Variable ist im Datensatz durch das Suffix „\_g#“ gekennzeichnet. Eine Übersicht aller für die 20. Sozialerhebung generierten Variablen sowie eine detaillierte Dokumentation der einzelnen Variablen mit Angabe ihrer jeweiligen Ausprägungen und Berechnungsvorschriften sind im Datensatzreport zu finden.

Generierte Variablen wurden im Datensatz – sofern möglich – nach der jeweiligen Ausgangsvariable positioniert. Wurde eine Variable aus verschiedenen Ausgangsvariablen generiert, wurde sie hinter jene Variable sortiert, welche ihr thematisch am ehesten entspricht. Falls eine thematische Zuordnung nicht möglich war, wurde die generierte Variable an das Ende des Datensatzes gestellt. Generierte Variablen, die aus Gründen der Anonymisierung bereitgestellt werden, wurden bis auf wenige Ausnahmen durch das FDZ erstellt.

## 6.5 Erstellung des Datensatzes

**[Datenstruktur und Dateiformat]** Der Datensatz enthält die Befragungsdaten sowie die zusätzlich generierten Variablen. Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen. Der Datensatz wird sowohl im Stata- als auch im SPSS-Format bereitgestellt (vgl. Abschnitt III).

## 6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels

**[Variablen- und Wertelabelvergabe]** Für Variablen- und Wertelabels wurden Formulierungen des Fragebogens übernommen oder prägnante Kurzformen dieser Formulierungen gewählt. Dabei basieren die Variablenlabels in der Regel auf dem entsprechenden Fragetext. Grundlage für die Wertelabels sind je nach Fragetyp die Texte der Antwortoptionen bzw. eine Kombination der Texte von Frage und Antwortoption. Bei generierten Variablen, denen bestimmte Klassifikationen zugrunde liegen, wurden für die Wertelabels die Bezeichnungen der Schlüssel der Klassifikation wortgetreu übernommen. Die Variablen- und Wertelabels liegen auf Deutsch und auf Englisch vor. Im SPSS-Format existiert für jede Sprache ein eigener Datensatz. Im Stata-Format wurden zweisprachige Labels im gleichen Datensatz hinterlegt.

**[Variablenbenennung]** Für die Variablenbenennung wurde im FDZ-DZHW eine einheitliche Benennungssystematik erstellt. Mit Ausnahme der Identifikatorvariablen werden die Variablennamen nach einem Präfix-Stamm-Suffix-Schema gebildet. Die Variablennamen liefern Metainformationen zur entsprechenden Variable. Dies erleichtert die automatisierte Verarbeitung. Das Präfix der Variable enthält bei Längsschnittbefragungen mit mehr als einem Befragungszeitpunkt (Panel) die Wellenkennung anhand eines Buchstabens. Da es sich bei der

20. Sozialerhebung um eine Querschnittsbefragung handelt, entfällt hier das Präfix. Im Stamm geht der Themenbereich, dem die Variable zugeordnet ist, aus einem dreistelligen englischen Buchstabenkürzel hervor. Tabelle 3 stellt die verschiedenen Themenbereiche der 20. Sozialerhebung sowie das zugehörige Kürzel für den Stamm des Variablennamens dar. Das anhand eines Unterstrichs vom Stamm abgetrennte Suffix enthält verschiedene Zusatzinformationen, wie die Kenntlichmachung von generierten Variablen, Panelvariablen sowie verschiedenen Datenzugangswegen.

Detaillierte Informationen zur Variablenbenennung in der 20. Sozialerhebung befinden sich im Datensatzreport.

**Tabelle 3: Themengebiete im Variablennamen der 20. Sozialerhebung (2012)**

Themengebiete-Kürzel	Themengebiet (englisch)	Themengebiet (deutsch)
<b>stu</b>	studies	Studium
<b>tim</b>	time usage (studies/job)	Zeitaufwand (Studium/Erwerbstätigkeit)
<b>ped</b>	prior education	Vorbildung
<b>fin</b>	financing (of living during studies)	Finanzierung (des Lebensunterhalts während des Studiums)
<b>baf</b>	BAföG (German Federal Grant on Training and Education Promotion)	BAföG (Bundesausbildungsförderungsgesetz)
<b>job</b>	job (during studies)	Erwerbstätigkeit (im Studium)
<b>liv</b>	living (accommodation)	Wohnform
<b>nut</b>	nutrition	Ernährung
<b>dem</b>	demographic information	Demographische Daten
<b>hea</b>	health	Gesundheit
<b>par</b>	parents	Eltern
<b>abr</b>	experiences abroad	Auslandserfahrungen
<b>lan</b>	language (skills)	Sprach(kenntnisse)
<b>ski</b>	(computer and internet) skills	Fähigkeiten (im Umgang Computer- und Internet)
<b>ciu</b>	computer and internet use	Computer- und Internetnutzung

## 6.7 Codierung fehlender Werte

Zur Codierung fehlender Werte wurde im FDZ-DZHW eine übergreifende Systematik erstellt, um über verschiedene Datensätze des DZHW hinweg eine einheitliche Missingcodierung gewährleisten zu können. Fehlende Angaben werden dabei durch dreistellige negative Werte codiert. Tabelle 4 stellt die verwendete Missingsystematik dar. Die in der 20. Sozialerhebung verwendeten Missingcodierungen sind hervorgehoben.

Sie lassen sich vier verschiedenen Gruppen zuordnen. In den ersten beiden Gruppen wird zwischen fehlenden Werten aufgrund von Nicht-Beantwortung von Fragen seitens der Befragten (Nonresponse) und fehlenden Werten aufgrund der Filterführung bzw. für Befragte nicht relevanten Fragen unterschieden (nicht zutreffend). Die dritte Gruppe beinhaltet Missingcodierungen, die durch das Primärforschungsprojekt oder das FDZ im Zuge der Datenaufberei-

ung vergeben wurden (editierter fehlender Wert). Zu dieser Gruppe gehört auch die Codierung für fehlende Werte, die aufgrund von Anonymisierungsmaßnahmen (vgl. Kapitel 8) für bestimmte Variablen vergeben wurde. Die vierte Gruppe umfasst spezielle Missingcodierungen, die im Rahmen der Erstellung eines konkreten Datensatzes nur für einzelne Items vergeben wurden. Im Datensatz der 20. Sozialerhebung gibt es keine solchen Item-spezifischen fehlenden Werte.

Tabelle 4: Systematik fehlender Werte des FDZ-DZHW

Wertebereich	Code	Wertelabel
<b>-999 bis -990: Nonresponse</b>	-999	weiß nicht
	<b>-998</b>	<b>keine Angabe</b>
	-997	keine Angabe (Antwortkategorie)
	-996	Interviewabbruch
	-995	keine Teilnahme (Panel)
	-994	verweigert
<b>-989 bis -970: Nicht zutreffend</b>	<b>-989</b>	<b>filterbedingt fehlend</b>
	<b>-988</b>	<b>trifft nicht zu</b>
	-987	designbedingt fehlend (Fragebogensplit)
	-986	designbedingt fehlend (Welle) <sup>a</sup>
	-985	designbedingt fehlend (Kohorte) <sup>b</sup>
<b>-969 bis -950: Editierter fehlender Wert</b>	-969	unbekannter fehlender Wert <sup>c</sup>
	-968	unplausibler Wert <sup>d</sup>
	<b>-967</b>	<b>anonymisiert</b>
	<b>-966</b>	<b>nicht bestimmbar<sup>e</sup></b>
	-965	ungültige Mehrfachnennung
<b>-949 bis -930: Item-spezifische fehlende Werte</b>	<i>(wurden in der 20. Sozialerhebung nicht vergeben)</i>	
<b>-929 bis -920: Andere fehlende Werte</b>	-929	Datenverlust

a Dieser Wert ist nur für Episoden-Datensätze im Long-Format vergeben.

b Dieser Wert wird nur in gepoolten Datensätzen vergeben.

c Dieser Wert wird vergeben, wenn keinerlei Ursache rekonstruiert werden kann.

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

e 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.

## 7 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen der Stichprobe aufgrund des Stichprobendesigns (vgl. Kapitel 3) und im Vergleich zur definierten Grundgesamtheit. Im Folgenden erfolgt zunächst eine allgemeine Einführung in die Vorgehensweise und eine Darstellung der erstellten Gewichte. Im Anschluss wird die Gewichtungsprozedur im Detail beschrieben.

### 7.1 Vorgehen und Anwendungshinweise

**[Ursachen für die Verzerrungen der Stichproben]** Maßgeblich für die Verzerrungen von Stichproben sind zwei Prozesse:

- **Designbedingte Verzerrung:** Disproportionalitäten werden bewusst erzeugt, um in bestimmten relevanten Subgruppen die Fallzahlen zu erhöhen (vgl. Kapitel 3).
- **Verzerrung durch Nonresponse:** Ausfallprozesse (z.B. Nichtteilnahmen, fehlende Erreichbarkeit, Verlust auf dem Postweg) führen zu einem verringerten Rücklauf und somit zu einer Differenz zwischen Brutto- und Nettostichprobe (vgl. Kapitel 5). Wenn diese Ausfallsprozesse unsystematisch sind (Missing Completely at Random), können sie ignoriert werden.<sup>39</sup> Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Missing Not at Random), der einer Modellierung bedarf.<sup>40</sup>

**[Konzeptuelles Vorgehen]** Im Zuge einer Gewichtungsprozedur sollten idealerweise zunächst designbedingte Disproportionalitäten ausgeglichen werden. Die hierfür benötigten *Designgewichte* ergeben sich bei zufallsgesteuerten Auswahlverfahren direkt aus dem Stichprobendesign. Im Anschluss sollte eine Adjustierung der Designgewichte mit Hilfe von *Nonresponsegewichten* im Quer- und Längsschnitt erfolgen, die auf der Grundlage von Informationen über Teilnehmer(innen) und Nichtteilnehmer(innen) auf Individualebene erzeugt werden. In einem letzten Schritt können die nonresponse-adjustierten Designgewichte anhand von Merkmalsverteilungen aus der Grundgesamtheit kalibriert werden (*Kalibrierung*).

Aufgrund des Stichprobendesigns (vgl. Kapitel 3) der 20. Sozialerhebung wird in einem ersten Schritt ein Designgewichte gebildet, um die ungleichen Inklusionswahrscheinlichkeiten auszugleichen. Da auf individueller Ebene keine Informationen zu Nichtteilnehmer(inne)n vorliegen, kann keine Nonresponse-Adjustierung des Designgewichts auf Individualebene erfolgen. Das Designgewicht wird somit in einem zweiten Schritt anhand einer Merkmalsverteilung der Grundgesamtheit kalibriert. Da hier Informationen über Teilnehmer(innen) und Nichtteilnehmer(innen) auf aggregierter Ebene vorliegen, erfolgt hier zugleich eine Nonresponse-Adjustierung.

<sup>39</sup> Insofern die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

<sup>40</sup> Siehe grundlegend zu den unterschiedlichen Formen von Ausfallprozessen Rubin (1976).

**Tabelle 5: Bereitgestellte Gewichte zur 20. Sozialerhebung (2012)**

Variablenname	Beschreibung
wgt01	Gewicht für Analysen auf Bundesebene
wgt02	Gewicht für Analysen auf Regionalebene (Nord, Süd, Ost, West)
wgt03	Gewicht für Analysen auf der Ebene Ost- versus Westdeutschland

**[Hinweise zur Anwendung der Gewichte]** Bei den erstellten Gewichten handelt es sich um *probability weights*, die in Stata mit Hilfe .ado-spezifischer Optionen berücksichtigt werden können.<sup>41</sup> Das Gewicht wgt01 ist für Auswertungen auf Bundesebene vorgesehen. Das Gewicht wgt02 bietet Gewichte für die Regionen Nord-, Süd-, Ost- und Westdeutschland. Das Gewicht wgt03 enthält Gewichte für die Regionen Ostdeutschland (neue Bundesländer inklusive Berlin) und Westdeutschland (alte Bundesländer ohne Berlin). Die Gewichte wgt02 und wgt03 dürfen jeweils nur in Auswertungen bezüglich *einer* der jeweiligen Regionen verwendet werden.<sup>42</sup> Grundlegend ist zu beachten, dass Gewichte nur dann sinnvolle Korrekturgrößen darstellen, wenn das verwendete Analysemodell die zur Gewichtung herangezogenen Variablen enthält oder mit diesen in einem Zusammenhang steht. Aus diesem Grund müssen Gewichte immer mit Fokus auf die analysierte Fragestellung verwendet werden. Im Folgenden wird die Vorgehensweise bei der Erstellung des Gewichtes näher dargestellt.

## 7.2 Gewichtung des Datensatzes

**[Designgewichtung]** Wie in Kapitel 3 beschrieben, sind Studierende einiger Hochschulen aufgrund des Stichprobendesigns überrepräsentiert. Folglich weisen diese Individuen eine höhere Auswahlwahrscheinlichkeit in die Stichprobe zu gelangen auf. Dies kann durch die Verwendung der inversen Auswahlwahrscheinlichkeit ( $\pi_i$ ) korrigiert werden. Das Designgewicht  $wgt\_d_i$  ist somit:

$$wgt\_d_i = \frac{1}{\pi_i}$$

Elemente, die mit höherer Wahrscheinlichkeit als andere in die Stichprobe eingehen, erhalten somit ein niedrigeres Gewicht und umgekehrt.

**[Kalibrierung der Designgewichte]** Wie beschrieben, war eine Nonresponse-Adjustierung der Designgewichte auf Individualebene nicht möglich. Es lagen jedoch folgende demographische Merkmale aus der Grundgesamtheit<sup>43</sup> vor, die zur Kalibrierung der Gewichte verwendet werden konnten: Region, Geschlecht, Fächergruppe, Hochschultyp, Personen mit deutscher

<sup>41</sup> Siehe hierzu die Stata-Hilfe (Befehl: *help weights*).

<sup>42</sup> Die Regionen können über die Variable *stu16\_g8r* (Nord-Süd-Ost-West-Aggregation für Remote-Desktop- und On-Site-Datensatz) bzw. *stu16\_g5* (Ost-West-Aggregation für Download-Datensatz) selektiert werden (z. B. mittels des Stata-Befehls *bysort*).

<sup>43</sup> Alle Informationen, die zur Kalibrierung der Designgewichte verwendet wurden, leiten sich aus Daten des Wintersemesters 2011/2012 des Statistischen Bundesamtes ab, da die aktuelle Statistik zum Zeitpunkt der Gewichtung noch nicht vorlag.

Staatangehörigkeit versus Bildungsinländer.<sup>44</sup> Die Kalibrierung erfolgte sowohl auf Bundesebene (wgt01) als auch gesondert für jede der vier Regionen (wgt02) sowie Ost- und Westdeutschland (wgt03). Die jeweiligen Regionalgewichte wurden in einer Variablen zusammengefasst (wgt02). Die Gewichte für Ost- und Westdeutschland wurden ebenfalls in einer Variablen zusammengefasst (wgt03).

Da die Merkmalsträger in der Grundgesamtheit ebenfalls Informationen über Nichtteilnehmer(innen) enthielten, erfolgte durch die Verwendung der Redressmentgewichte zusätzlich eine Art Nonresponse-Adjustierung im Hinblick auf die verwendeten Merkmale. Die Kalibrierung wurde mittels des Raking-Algorithmus<sup>45</sup> durchgeführt.

**[Trimmung der Gewichte]** Die initial berechneten Gewichte wiesen einen kleinen Teil an Gewichtungsfaktoren auf, die Ausreißer darstellten. Um diese zu beseitigen, wurden alle Gewichte einer Trimmung nach Potter (1990) (vgl. auch Valliant, Dever & Kreuter, 2013, S. 388f.) unterzogen. Dem Verfahren liegt die Annahme zugrunde, dass die Gewichte einer Wahrscheinlichkeitsverteilung (Beta-Verteilung) folgen. All jene Gewichte, die über dem 99-Prozent-Quantil liegen, werden auf diese Grenze trunziert. Der Überschuss jenseits der Trunkierung wird im Folgenden unter den verbleibenden Gewichten verteilt.

---

<sup>44</sup> Die Gewichtung wurde entlang folgender Ausprägungen durchgeführt: Geschlecht: weiblich versus männlich; Region: Nord (Bremen, Hamburg, Niedersachsen, Schleswig-Holstein), Süd (Baden-Württemberg, Bayern), Ost (Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, Thüringen), West (Hessen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland); Hochschultyp: Universität versus Fachhochschule (inklusive Pädagogische Hochschulen, Theologische Hochschulen, Kunst- und Musikhochschulen); Fächergruppe: entsprechend Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2011/2012 und SoSe 2012).

<sup>45</sup> Raking ist auch unter dem Begriff des iterative proportional fitting (ipf) bekannt (Kolenikov (2014)).



## 8 Anonymisierung

**[Datenschutzrechtlicher Rahmen]** Für personenbezogene Daten, die in freiwilligen Befragungen durch das DZHW erhoben werden, gilt das Bundesdatenschutzgesetz (BDSG).<sup>46</sup> Danach dürfen personenbezogene Daten, die im Rahmen wissenschaftlicher Forschung erhoben worden sind, ausschließlich zum Zweck wissenschaftlicher Forschung verarbeitet oder genutzt werden (vgl. § 40 Abs. 1 BDSG). Darüber hinaus sind personenbezogene Daten im Forschungskontext zum Schutz der Befragten zu anonymisieren (vgl. § 40 Abs. 2 BDSG). Im BDSG wird der Vorgang der Anonymisierung definiert als „das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können“ (§ 3 Abs. 6 BDSG).<sup>47</sup> Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten entweder absolut zu anonymisieren, so dass kein Bezug zur Person mehr hergestellt werden kann, oder mindestens faktisch zu anonymisieren, so dass die Herstellung eines Personenbezugs mit einem unverhältnismäßig hohen Aufwand verbunden und daher die Wahrscheinlichkeit der Re-Identifikation einer Person minimal ist.

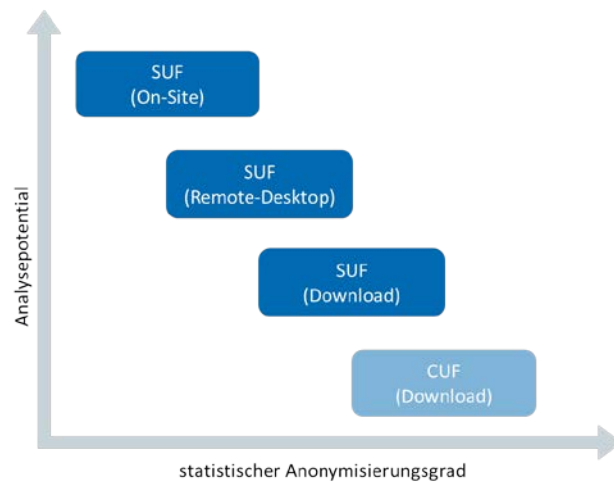
**[Datenzugang, Anonymisierungsgrad und Analysepotential]** Das FDZ-DZHW stellt für die 20. Sozialerhebung (2012) ein faktisch anonymisiertes SUF für die wissenschaftliche Sekundärnutzung und ein absolut anonymisiertes CUF für Lehr- und Übungszwecke zur Verfügung. Die Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen und technischen Zugriffsbeschränkungen sichergestellt. Je stärker der Datenzugang technisch kontrolliert wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten, desto weniger müssen die Daten mittels statistischer Maßnahmen um Informationen reduziert werden und desto größer bleibt ihr Analysepotential.

Während das CUF nach einer Registrierung direkt durch das FDZ-DZHW übermittelt wird, wird das SUF über drei verschiedene Zugangswege angeboten: Download, Remote-Desktop und On-Site (für weiterführende Informationen vgl. Abschnitt III). Für jeden Zugangsweg wird eine andere SUF-Variante bereitgestellt, die unterschiedlich stark anonymisiert worden ist und entsprechend weniger oder mehr Informationen umfasst. Abbildung 3 gibt einen Überblick über den jeweiligen Grad der statistischen Anonymisierung und dem damit verbundenen Analysepotential. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen in Abhängigkeit von Datenprodukt (SUF/CUF) und Zugangsweg erläutert.

<sup>46</sup> Das BDSG kommt zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (§ 2 Abs. 3 BDSG). Der Bund hält die absolute Mehrheit der Anteile der DZHW GmbH und das Institut erfüllt Aufgaben der öffentlichen Verwaltung des Bundes im weitesten Sinn. Zur Auslegung einzelner rechtlicher Aspekte wird ergänzend die Europäische Datenschutzrichtlinie hinzugezogen.

<sup>47</sup> Wenn die Einzelangaben **nicht mehr** einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können, handelt es sich um „absolute Anonymisierung“. Wenn die Einzelangaben nur mit einem **unverhältnismäßig großen Aufwand** an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmaren natürlichen Person zugeordnet werden können, spricht man von faktischer Anonymisierung.

**Abbildung 3: Datenzugangsweg, statistischer Anonymisierungsgrad und Analysepotential der Daten der 20. Sozialerhebung (2012)**



**[Statistische Anonymisierungsmaßnahmen]** Im Rahmen der Anonymisierung sind zunächst alle Informationen, mit denen sich Personen oder Institutionen direkt identifizieren lassen, zu löschen. Von diesen sogenannten *direkten Identifikatoren*, wie Namen, Adressen oder E-Mail-Adressen, wurde im Rahmen der 20. Sozialerhebung nur die letztgenannte erfasst, und diese auch nur dann, wenn Personen auf der letzten Fragebogenseite ihre Bereitschaft erklärten, am HISBUS-Panel<sup>48</sup> teilzunehmen. Die E-Mail-Adressen wurden unmittelbar nach Öffnung des Rückumschlags vom Fragebogen abgetrennt und in einem separaten Datensatz ohne die Möglichkeit des Rückbezugs zum Datensatz der 20. Sozialerhebung erfasst. Sie sind weder im CUF noch in den verschiedenen SUF-Varianten enthalten. Des Weiteren wurde, um einen Rückbezug auf die Originaldaten und Fragebögen zu verhindern, die Original-Identifikationsnummer aus dem Datensatz entfernt und durch eine neue, zufällig vergebene Identifikationsnummer ersetzt.

Anschließend wurden die *Quasi-Identifikatoren* bestimmt, also Informationen, die in Kombination oder durch die Anspielung externer Informationen geeignet sind, eine Person indirekt zu identifizieren.<sup>49</sup> Für die 20. Sozialerhebung wurden die folgenden Quasi-Identifikatoren identifiziert, die sowohl in externen Datenquellen<sup>50</sup> als auch in der 20. Sozialerhebung vorliegen: Name sowie Art und Ort der Hochschule, Studienfach, Abschlussart, Alter und Staatsangehörigkeit. Um eine eindeutige Zuordnung der Daten der 20. Sozialerhebung zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenprodukt bzw. Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 6). Beispielsweise ist das Merkmal „Studienfach“ im SUF

<sup>48</sup> Beim HISBUS-Panel handelt es sich um seit 2002 regelmäßig vom DZHW durchgeführte Online-Befragungen von Studierenden zu bildungspolitisch bedeutenden Themen. Die befragten Studierenden haben sich freiwillig bereit erklärt, am Panel teilzunehmen und werden aus anderen Befragungen des DZHW rekrutiert. Nähere Informationen zum HISBUS-Panel sind unter <http://www.hisbus.de> verfügbar.

<sup>49</sup> Dabei ist darauf hinzuweisen, dass die Identifikation einer Person bereits durch die Stichprobenauswahl erschwert wird, da eine Ungewissheit darüber besteht, ob eine befragte Person eine einzigartige Merkmalskombination in der Population aufweist.

<sup>50</sup> z. B. Studenten- und Prüfungsstatistik des Statistischen Bundesamtes, Alumninetzwerke der Hochschulen oder auch Berufsnetzwerke.

für die On-Site-Nutzung uneingeschränkt verfügbar. Im Remote-Desktop-SUF hingegen wird das Merkmal zu den Studienbereichen der amtlichen Statistik aggregiert. Im Download-SUF sowie im CUF sind nur die Fächergruppen der amtlichen Statistik verfügbar. Offene Angaben sind ebenfalls Quasi-Identifikatoren (vgl. Ebel, 2015, S. 3). Sie wurden teilweise bereits im Rahmen der Datenaufbereitung durch das Primärforschungsprojekt vercodet. Codierte offene Angaben werden im CUF sowie in allen SUF-Varianten zur Verfügung gestellt. Teilweise wurden jedoch – in Abhängigkeit von der Sensibilität der enthaltenen Informationen und vom Zugangsweg – die vom Primärforschungsprojekt vorgenommenen Codierungen zusätzlich aggregiert. Nicht codierte offene Angaben (vgl. Kapitel 6.2) wurden im CUF und in allen SUF-Varianten gelöscht.

Zuletzt wurde geprüft, ob in den Daten *sensible Informationen*, z. B. zur Gesundheit, sexuellen Orientierung oder zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (vgl. (Koberg, 2016, S. 694) und sind daher besonders schützenswert (vgl. §3 Abs. 9 BDSG, Art. 8 Abs. 1 und 2a EG-DSRL). In der 20. Sozialerhebung wurden Gesundheitsinformationen erhoben, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Daher wurden diese Antworten im CUF und allen SUF-Varianten gelöscht.

Zur Realisierung der absoluten Anonymität der Befragten im CUF wurde zusätzlich zu den vergleichsweise restriktiven statistischen Anonymisierungsmaßnahmen des Download-SUF eine 10-Prozent-Substichprobe der Download-SUF-Daten gezogen. Die nachfolgende Tabelle stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen je nach Datenform bzw. Zugangsweg dar.

**Tabelle 6: Maßnahmen der statistischen Anonymisierung der Daten der 20. Sozialerhebung (2012) nach Zugangsweg<sup>51</sup>**

<b>Merkmal</b>	<b>On-Site-SUF</b>	<b>Remote-Desktop-SUF</b>	<b>Download-SUF</b>	<b>Download-CUF (Substichprobe)</b>
Direkte Identifikatoren	Löschung	Löschung	Löschung	Löschung
Original-ID	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID	Löschung und Vergabe einer zufälligen ID
Studienfächer	Freigabe	Aggregation zu Studienbereichen <sup>a</sup>	Aggregation zu Fächergruppen <sup>a</sup>	Aggregation zu Fächergruppen <sup>a</sup>
Abschlussart	Freigabe	Freigabe	Aggregation <sup>b</sup>	Aggregation <sup>b</sup>
Semesterzahl bis Fach-/ Hochschulwechsel	Freigabe	Freigabe	1-10 einzeln ausgewiesen, ansonsten aggregiert (mehr als 10)	1-10 einzeln ausgewiesen, ansonsten aggregiert (mehr als 10)
Monate der Erwerbstätigkeit zwischen 1. Abschluss und Masterstudium	Freigabe	Freigabe	Aggregation: 1-6; 7-12; 13-24; 25-36; 37-60; 61-120; mehr als 120	Aggregation: 1-6; 7-12; 13-24; 25-36; 37-60; 61-120; mehr als 120
Hochschule	Aggregation zu Hochschulart	Aggregation zu Hochschulart	Aggregation zu Hochschulart <sup>c</sup>	Aggregation zu Hochschulart <sup>c</sup>
Hochschulort	Vier Bundesländer einzeln ausgewiesen; ansonsten Aggregation zu fünf Bundesländer-Gruppen	Vier Bundesländer einzeln ausgewiesen; ansonsten Aggregation zu fünf Bundesländer-Gruppen	Aggregation zu neuen bzw. alten Bundesländern	Aggregation zu neuen bzw. alten Bundesländern
Gründe für Unterbrechung des Studiums	Aggregation <sup>d</sup>	Aggregation <sup>d</sup>	Aggregation <sup>d</sup>	Aggregation <sup>d</sup>
Anzahl Unterbrechungssemester	Freigabe	Freigabe	1-4 einzeln ausgewiesen, ansonsten Aggregation: 5-10; mehr als 10	1-4 einzeln ausgewiesen, ansonsten Aggregation: 5-10; mehr als 10
Wartezeit bis Studienbeginn (in Monaten)	Freigabe	Freigabe	Aggregation: 0-12; 13-24; 25-36; 37-60; 61-120; mehr als 120	Aggregation: 0-12; 13-24; 25-36; 37-60; 61-120; mehr als 120
andere Finanzierungsquellen	Freigabe	Aggregation <sup>e</sup>	Aggregation zu einer Gruppe: „Quelle genannt“	Aggregation zu einer Gruppe: „Quelle genannt“
Förderungsform (BAföG)	Freigabe	Freigabe	Löschung	Löschung

<sup>51</sup> Detaillierte Informationen zu den anonymisierten Variablen sind dem Datensatzreport sowie dem Metadaten-suchsystem (<https://metadata.fdz.dzhw.eu>) zu entnehmen.

Gründe für keine Erwerbstätigkeit	Aggregation <sup>f</sup>	Aggregation <sup>f</sup>	Aggregation <sup>f</sup>	Aggregation <sup>f</sup>
Alter (in Jahren)	Freigabe	0 bis 39 einzeln ausgewiesen, ansonsten Aggregation: 40-44; 45-49;...; 70-74	0 bis 39 einzeln ausgewiesen, ansonsten Aggregation: 40-44; 45-49;...; 70-74	0 bis 39 einzeln ausgewiesen, ansonsten Aggregation: 40-44; 45-49;...; 70-74
Anzahl der Geschwister in Ausbildung	Freigabe	1, 2 und 3 einzeln ausgewiesen, ansonsten Aggregation: 4 oder mehr	1 und 2 einzeln ausgewiesen, ansonsten Aggregation: 3 oder mehr	1 und 2 einzeln ausgewiesen, ansonsten Aggregation: 3 oder mehr
Anzahl der Kinder	Freigabe	1 und 2 einzeln ausgewiesen, ansonsten Aggregation: mehr als 2	1 und 2 einzeln ausgewiesen, ansonsten Aggregation: mehr als 2	1 und 2 einzeln ausgewiesen, ansonsten Aggregation: mehr als 2
Alter des jüngsten Kindes (in Jahren)	Freigabe	Aggregation: 1-2; 3-5; 6-11; 12-17; 18 und älter	Aggregation: 1-5; 6-17; 18 und älter	Aggregation: 1-5; 6-17; 18 und älter
Beruf der Eltern (sowie Erläuterungen dazu)	Löschung	Löschung	Löschung	Löschung
andere als deutsche Staatsangehörigkeit	Freigabe	5 Staaten einzeln ausgewiesen, ansonsten Aggregation zu Weltregionen	Aggregation zu Weltregionen	Aggregation zu Weltregionen
vorherige Staatsangehörigkeit	Freigabe	5 Staaten einzeln ausgewiesen, ansonsten Aggregation zu Weltregionen	Aggregation zu Weltregionen	Aggregation zu Weltregionen
andere als deutsche Staatsangehörigkeit der Eltern	Freigabe	5 Staaten einzeln ausgewiesen, ansonsten Aggregation zu Weltregionen	Aggregation zu Weltregionen	Aggregation zu Weltregionen
Studienbezogener Auslandsaufenthalt (1./2. Land)	Freigabe	Freigabe	5 Staaten einzeln ausgewiesen, ansonsten Aggregation zu Weltregionen	5 Staaten einzeln ausgewiesen, ansonsten Aggregation zu Weltregionen
Studienbezogener Auslandsaufenthalt (3./4. Land)	Freigabe	Freigabe	Aggregation zu Weltregionen	Aggregation zu Weltregionen
Gesundheitliche Beeinträchtigung (Frage 41)	Löschung	Löschung	Löschung	Löschung

<sup>a</sup> nach Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2011/2012 und SoSe 2012 von Destatis

- <sup>b</sup> "kirchliche Prüfung" und "anderer Abschluss" zu "sonstiger Abschluss (kirchliche Prüfung, anderer Abschluss)" zusammengefasst
- <sup>c</sup> nur Unterscheidung von Fachhochschule und Universität (inklusive Pädagogische Hochschulen, Theologische Hochschulen, Kunst- und Musikhochschulen)
- <sup>d</sup> Zusammenfassung der Gründe „Betreuung von pflegebedürftigen Angehörigen“, „akute gesundheitliche Probleme“ und „chronische Krankheit/Behinderung“ mit „sonstiger Grund“ zu einer Gruppe
- <sup>e</sup> einzeln ausgewiesen: „Leistungen für eigene Kinder/wegen eigener Kinder“; ansonsten Aggregation: „sonstige Sozialleistungen“ und „sonstige Quellen“
- <sup>f</sup> Zusammenfassung der Gründe „Behinderung/gesundheitliche Beeinträchtigung“ und „Pflege von Angehörigen“ mit „Kindererziehung“ zu einer Gruppe

## 9 Literaturverzeichnis

- Apolinarski, B. (2013). *Die wirtschaftliche und soziale Lage der Studierenden in München. Regionalauswertung der 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung* (Studentenwerk München, Hrsg.). München: München. Verfügbar unter [http://www.studentenwerk-muenchen.de/fileadmin/studentenwerk-muenchen/publikationen/broschueren/stwm\\_Sozialbericht\\_20140326\\_web.pdf](http://www.studentenwerk-muenchen.de/fileadmin/studentenwerk-muenchen/publikationen/broschueren/stwm_Sozialbericht_20140326_web.pdf)
- Apolinarski, B. (2014). *Die wirtschaftliche und soziale Lage der Studierenden in Köln 2012. Regionalauswertung der 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung* (Kölner Studentenwerk, Hrsg.), Köln. Verfügbar unter [http://kstw.de/images/stories/presse/regionalbericht\\_kstw\\_12%2002%2014\\_fin.pdf](http://kstw.de/images/stories/presse/regionalbericht_kstw_12%2002%2014_fin.pdf)
- Apolinarski, B., Buck, D., Kandulla, M., Middendorff, E., Naumann, H. & Poskowsky, J. (2014). *Daten- und Methodenbericht zum Scientific Use File der 20. Sozialerhebung. Version 1.0.0* (Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW), Hrsg.) (Dokumentation). Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
- Apolinarski, B. & Poskowsky, J. (2013). *Ausländische Studierende in Deutschland 2012. Ergebnisse der 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt vom Deutschen Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)* (BMBF, Hrsg.), Bonn. Verfügbar unter [http://www.sozialerhebung.de/download/20/soz20\\_auslaenderbericht.pdf](http://www.sozialerhebung.de/download/20/soz20_auslaenderbericht.pdf)
- Ebel, T. (2015). *Empfehlungen zur Anonymisierung quantitativer Daten*. Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften.
- Häder, M. (2015). *Empirische Sozialforschung. Eine Einführung* (3. Aufl.). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-531-19675-6
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). *FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum* (Nr. 06). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Isserstedt, W., Middendorff, E., Kandulla, M., Borchert, L. & Leszczensky, M. (2010). *Die wirtschaftliche und soziale Lage der Studierenden in der Bundesrepublik Deutschland 2009. 19. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch HIS Hochschul-Informationssystem* (Bundesministerium für Bildung und Forschung (BMBF), Hrsg.) (Wissenschaft Ideen zünden!), Bonn
- Kandulla, M. (2013). *Die wirtschaftliche und soziale Lage der Studierenden in Berlin. Regionalauswertung Berlin der 20. Sozialerhebung des Deutschen Studentenwerks*. (Studentenwerk Berlin., Hrsg.), München. Verfügbar unter <http://www.studentenwerk-berlin.de/studentenwerk/dokumente/65%20%7C%20Soziale%20Lage%20Studierende.pdf>
- Kandulla, M. (2014). *Die wirtschaftliche und soziale Lage der Studierenden in Nordrhein-Westfalen. Sonderauswertung der Arbeitsgemeinschaft der Studentenwerke NRW zur 20. Sozialerhebung des Deutschen Studentenwerks im Jahre 2012* (Arbeitsgemeinschaft der Studentenwerke NRW, Hrsg.), Bielefeld. Verfügbar unter [http://www.studentenwerke-nrw.de/CMS/images/stories/STWeData/documents/PDFs/Regionalbericht\\_ARGE\\_STWNRW\\_Gesamtbericht\\_06\\_02\\_14.pdf](http://www.studentenwerke-nrw.de/CMS/images/stories/STWeData/documents/PDFs/Regionalbericht_ARGE_STWNRW_Gesamtbericht_06_02_14.pdf)
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer & J. Skopek (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
- Kolenikov, S. (2014). Calibrating survey data using iterative proportional fitting (raking). *The Stata Journal*, 14 (1), 22–59.
- Lane, J., Heus, P. & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy*, 1 (1), 2–16.
- Middendorff, E. (2016). *Die Sozialerhebungen des Deutschen Studentenwerks 1951 – 2016. Ein historischer Überblick über Akteure, Wellen und Themen*. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
- Middendorff, E., Apolinarski, B., Poskowsky, J., Kandulla, M. & Netz, N. (2013). *Die wirtschaftliche und soziale Lage der Studierenden in Deutschland 2012. 20. Sozialerhebung des Deutschen Studentenwerks durchgeführt durch das HIS-Institut für Hochschulforschung* (BMBF, Hrsg.), Bonn. Verfügbar unter [http://www.sozialerhebung.de/download/20/soz20\\_hauptbericht\\_gesamt.pdf](http://www.sozialerhebung.de/download/20/soz20_hauptbericht_gesamt.pdf)
- Middendorff, E. & Poskowsky, J. (2014). *Die wirtschaftliche und soziale Lage der Studierenden in Hannover. Sonderauswertung der 20. Sozialerhebung des Deutschen Studentenwerks 2012* (Studentenwerk Hannover, Hrsg.), Hannover. Verfügbar unter <http://www.studentenwerk-hannover.de/fileadmin/daten/pdf/allgemein/StwH-Sozialerhebung-2012.pdf>

- Netz, N. (2014). Der Zugang zu studienbezogenen Auslandsaufenthalten aus Sicht von Hochschulpolitik und Hochschulforschung: Eine Bestandsaufnahme. In U. Banscherus, M. Bülow-Schramm, K. Himpele, S. Staack & S. Winter (Hrsg.), *Übergänge im Spannungsfeld von Expansion und Exklusion. Eine Analyse der Schnittstellen im deutschen Hochschulsystem* (GEW Materialien aus Hochschule und Forschung, Bd. 121, S. 81–97). Bielefeld: Bertelsmann. doi:10.3278/6001596w
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63 (2), 581–592.
- Staneva, M. *Studieren und Arbeiten. Die Bedeutung der studentischen Erwerbstätigkeit für den Studienerfolg und den Übergang in den Arbeitsmarkt*. (DIW Roundup Nr. 70). Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).