

Daten- und Methodenbericht  
Juli 2021

Sören Isleib | Henrike Schmidtchen | Anne Weber

# StuMa: Ursachen der Studienaufgabe im Masterstudium

Daten- und Methodenbericht zur DZHW-Befragung  
StuMa 2020

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/de/>



#### **Autor\*innen**

Sören Isleib

Telefon +49 (0)511 450670-195

E-Mail: [isleib@dzhw.eu](mailto:isleib@dzhw.eu)

Henrike Schmidtchen

Telefon +49 (0)511 450670-419

E-Mail: [schmidtchen@dzhw.eu](mailto:schmidtchen@dzhw.eu)

Anne Weber

Telefon +49 (0)511 450670-488

E-Mail: [a.weber@dzhw.eu](mailto:a.weber@dzhw.eu)

#### **Impressum**

##### **Herausgeber**

Deutsches Zentrum für Hochschul- und  
Wissenschaftsforschung GmbH (DZHW)

Lange Laube 12 | 30159 Hannover | [www.dzhw.eu](http://www.dzhw.eu)

Postfach 2920 | 30029 Hannover

Tel.: +49 511 450670-0 | Fax: +49 511 450670-960

##### **Geschäftsführerinnen:**

Prof. Dr. Monika Jungbauer-Gans

Karen Schlüter

Vorsitzender des Aufsichtsrats:

Ministerialdirigent Peter Greisler

##### **Registergericht:**

Amtsgericht Hannover | B 210251

Umsatzsteuer-Identifikationsnummer:

DE291239300

Juli 2021

# Inhaltsverzeichnis

1	Datennutzungshinweise	2
2	Inhalt und Anlage der Studie	4
3	Methoden	5
3.1	Grundgesamtheit und Stichprobe.....	5
3.2	Feldphase und Rücklauf .....	5
4	Datenaufbereitung	7
5	Anonymisierung	9
6	Literatur	15

## Tabellen-/Abbildungsverzeichnis

Abbildung 1:	Datenzugangswege und Analysepotential.....	3
Tabelle 1:	Brutto-/Nettostichprobengröße und Rücklaufquote.....	6
Tabelle 2:	Verwendete Missingsystematik.....	8
Tabelle 3:	Maßnahmen der statistischen Anonymisierung der Daten von StuMa 2020 .....	11

# 1 Datennutzungshinweise

**[Voraussetzungen der Datennutzung]** Die Daten der DZHW-Befragung StuMa 2020 werden durch das FDZ des DZHW entsprechend der europäischen Datenschutzgrundverordnung (EU-DSGVO) anonymisiert bereitgestellt und ausschließlich zur wissenschaftlichen Nutzung freigegeben.<sup>1</sup> Das FDZ bietet sowohl ein *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung als auch ein *Campus Use File* (CUF) für Lehr- und Übungszwecke an. Beide können über die Website des FDZ (<https://fdz.dzhw.eu>) beantragt werden. Für die Nutzung von SUF wird ein Datennutzungsvertrag abgeschlossen.

**[Datenzugang]** Das CUF von StuMa 2020 wird via Download bereitgestellt und kann am lokalen Computer genutzt werden. Das SUF wird über zwei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind.

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

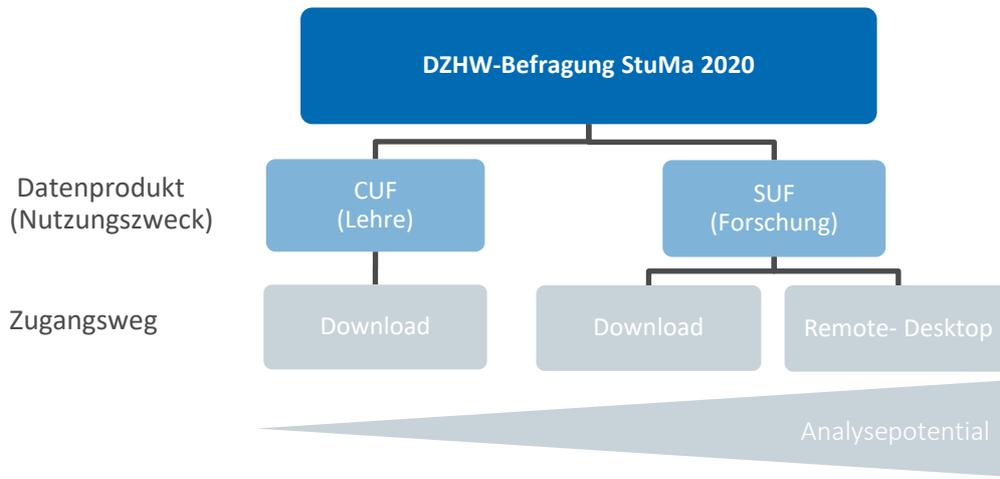
Die bereitgestellten Daten weisen je nach Zugangsweg einen unterschiedlich hohen Informationsgehalt auf und unterscheiden sich damit hinsichtlich ihres Analysepotentials (siehe hierzu Abbildung 1). Dabei gilt: Je stärker der Datenzugriff der Nutzer\*innen durch technische und organisatorische Maßnahmen kontrolliert wird, desto mehr Informationen können für die Datennutzer\*innen bereitgestellt werden.<sup>2</sup> Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

---

<sup>1</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 des DZHW hat diesen Ansatz an die Anforderungen der eigenen Datenbestände angepasst und nutzt vier Kategorien von Maßnahmen zur Sicherstellung des Datenschutzes, die in unterschiedlicher Weise kombiniert werden können: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen.

<sup>2</sup> Zu den verschiedenen Anonymisierungsgraden und Analysepotentials des CUF und der beiden SUF-Varianten vgl. Kapitel 5.

Abbildung 1: Datenzugangswege und Analysepotential



**[Gebühren der Datenbereitstellung]** CUF und SUF werden derzeit (Stand: Juli 2021) 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>3</sup> einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzer\*innen Kenntnis von Sicherheitslücken hinsichtlich Datenschutz bzw. Datensicherheit erlangen, müssen diese dem FDZ unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** SUF dürfen nur durch die Personen genutzt werden, die einen Datennutzungsvertrag abgeschlossen haben. CUF dürfen ausschließlich im Rahmen von dem FDZ gemeldeten Lehrveranstaltungen weitergegeben werden.
- **Löschungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel 1,5 Jahre) von jeglichen Rechnern, Servern und Datenträgern zu löschen. Ebenso müssen alle Sicherungskopien, modifizierten Datensätze (z. B. Arbeits-, Auszugs- oder Hilfsdateien) sowie Ausdrucke vernichtet werden.
- **Bereitstellung/Meldung von Publikationen:** Jede Art von wissenschaftlichen Arbeiten (z.B. Publikationen, Abschlussarbeiten, Dissertationen), die aus der Arbeit mit Daten des FDZ hervorgehen, ist dem FDZ anzuzeigen. Dabei ist dem FDZ eine elektronische Version der Druckfassung zur Verfügung zu stellen.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut der Vorgaben des FDZ zitiert werden.

<sup>3</sup> Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

## 2 Inhalt und Anlage der Studie

**[Inhalt der Studie]** Die Studie “StuMa” beschäftigt sich mit den Ursachen des Studienabbruchs im Masterstudium und dem weiteren beruflichen Verlauf nach einem Studienausstieg aus dem Masterstudium. Während für grundständige Studiengänge einige Daten vorliegen, die eine Analyse möglicher Ursachen des Studienabbruchs erlauben, existiert für das Masterstudium keine geeignete Datenbasis. Bedingt durch vergleichsweise geringe Studienabbruchquoten im Masterstudium, wurde der Studienabbruch in dieser Studienphase bislang auch noch nicht als Forschungsgegenstand in der Hochschulforschung wahrgenommen. StuMa ist die erste bundesweite Befragung, die sich ausschließlich an Masterexmatrikulierte richtet und eine Datenbasis generiert, um die Studiensituation und die Abbruchursachen im Masterstudium zu ermitteln sowie darüber hinaus den beruflichen Verbleib der Exmatrikulierten zu analysieren.

Für eine Analyse der Studienaufgabe im Masterstudium stellten sich in der Untersuchung des DZHW folgende zentrale Forschungsfragen:

- Was sind die Ursachen der Studienaufgabe im Masterstudium? Welche zentralen Faktoren beeinflussen den Prozess der Studienaufgabe?
- Inwiefern besteht Bedarf zu Prävention und Interventionen für eine Vermeidung von Studienaufgaben im Masterstudium? Welche Präventions- und Interventionsmöglichkeiten lassen sich ableiten?
- Wie entwickelt sich der berufliche Verbleib von Exmatrikulierten, die ihr Masterstudium aufgegeben haben?

**[Anlage der Studie]** Um diesen Fragen nachzugehen, wurde eine Online-Befragung mit Exmatrikulierten (mit und ohne Abschluss) aus Masterstudiengängen an staatlichen deutschen Hochschulen des Wintersemesters 2018/19 und des Sommersemesters 2019 durchgeführt.

**[Themen der Erhebung]** Neben allgemeinen Angaben, die das Masterstudium oder das zum Master berechtigende Studium betreffen (bspw. Hochschule, Studienfach, Semester), wurden auch Themen wie die Aufnahme des Studiums oder die aktuelle Studiensituation in der Befragung erfasst. Außerdem wird die Erwerbstätigkeit während des Studiums sowie der Vergleich zwischen der aktuellen und zukünftigen Tätigkeit näher beleuchtet. Ein weiterer Teil der Fragen beschäftigt sich mit dem bisherigen Bildungsweg und den soziodemographischen Merkmalen der Befragten.

**[Sekundär genutzte Erhebungsinstrumentbestandteile]** Im Erhebungsinstrument wurde neben den vorgestellten Themen auch die Persönlichkeit der Befragten erfasst. Hierzu wurden die Big Five: Neurotizismus, Extraversion, Offenheit, Gewissenhaftigkeit und Verträglichkeit mittels 16 Items erhoben. 15 dieser Items beziehen sich auf das Big Five Inventory-SOEP (BFI-S) (Schupp & Gerlitz, 2014). Erweitert wurde dieses Instrument durch eine Variable (selwiss) aus dem Studierendenberechtigten-Panel.

Im Folgenden wird über die methodischen Details der Befragung (Grundgesamtheit, Stichprobenziehung, Feldphase und Rücklauf) sowie über die Datenaufbereitung und Anonymisierung für das Scientific Use File (SUF) und das Campus Use File (CUF) informiert.

## 3 Methoden

### 3.1 Grundgesamtheit und Stichprobe

**[Grundgesamtheit]** Die Grundgesamtheit der Befragung sind Exmatrikulierte (mit und ohne Abschluss) aus Masterstudiengängen an staatlichen deutschen Hochschulen des Wintersemesters 2018/19 und des Sommersemesters 2019. Exmatrikulierte von Bundeswehrhochschulen, Verwaltungsfachhochschulen, Berufsakademien und Fernhochschulen, privaten Hochschulen sowie weiterer Hochschulen, die sich ausschließlich an Berufstätige richten, waren dabei von vornherein ausgenommen.

**[Stichprobenverfahren]** Die Individuen wurden über die Hochschulen rekrutiert. Es wurde zunächst eine Hochschulstichprobe gezogen. Dazu wurden auf Basis von Daten aus zurückliegenden Exmatrikulationssemestern sowie nach regionalen Gesichtspunkten Listen der zu ziehenden Hochschulen erstellt. Ziel war es a), Hochschulen zu gewinnen, bei denen auf Basis älterer Daten zahlenmäßig starke Jahrgänge in den interessierenden Exmatrikulationssemestern zu erwarten waren, so dass pro teilnehmender Hochschule der Zugang zu einer großen Zahl an potentiellen Befragten ermöglicht wurde. Ziel war es b), eine möglichst gleichmäßige regionale Abdeckung der verschiedenen Hochschularten (Universitäten, Technische Universitäten, Fachhochschulen) zu gewährleisten. Das heißt: Für jede Region Ost, Süd, West und Nord<sup>4</sup> wurde je eine Liste pro Hochschulart erstellt und absteigend nach zu erwartender Zahl an Exmatrikulierten sortiert. Hiernach wurden die Hochschulen für die Teilnahme rekrutiert. Ausnahmen bestanden dann, wenn Hochschulen ihre Teilnahme bereits im Vorfeld des Projekts signalisiert oder diese Studie wesentlich mit angeregt haben. Sie wurden dann unabhängig von zu erwartenden Fallzahlen gesetzt. Ebenso wurden bestimmte Hochschulen, die sich in der Vergangenheit DZHW-Befragungen entzogen haben, nicht angesprochen. Es konnten somit 28 Hochschulen für die Teilnahme gewonnen werden.

### 3.2 Feldphase und Rücklauf

Die Befragung erfolgte online. Die Einladung der Exmatrikulierten erfolgte per E-Mail durch die Hochschulen, denen das DZHW im Vorfeld alle notwendigen Materialien (Befragungslinks, Datenschutzhinweise, Anschreiben) zur Verfügung gestellt hat. Im Falle von zwei Hochschulen war die Einladung per E-Mail aufgrund fehlender E-Mailadressen der Exmatrikulierten nicht möglich. In diesen Fällen erfolgte die Einladung schriftlich-postalisch. Hier war ein Befragungslink enthalten, der per Hand in die Browser-Zeile eingegeben werden musste. Bei den beiden betreffenden Hochschulen handelt es sich um vergleichsweise kleine Hochschulen mit geringer Anzahl an Exmatrikulierten aus Masterstudiengängen (jeweils ca. 300), so dass für die Befragung und den Rücklauf insgesamt nur ein geringer Effekt durch dieses Einladungsverfahren zu erwarten war. Die Feldphase dauerte vom 06.01.2020 bis 10.04.2020.

**[Rücklaufsteigernde Maßnahmen]** Im Zuge der Befragung wurden rücklaufsteigernde Maßnahmen eingesetzt. Zum einen wurden nach individueller Absprache mit den Hochschulen bis zu drei Erinne-

<sup>4</sup> Ost: Brandenburg, Berlin, Sachsen-Anhalt, Sachsen, Thüringen; Süd: Bayern, Baden-Württemberg; West: Hessen, Rheinland-Pfalz, Saarland, Nordrhein-Westfalen; Nord: Niedersachsen, Bremen, Hamburg, Schleswig-Holstein, Mecklenburg-Vorpommern

rungsmails an die Befragten durch die Hochschulen versandt. Eine Hochschule wünschte dabei keine Erinnerung, 12 Hochschulen verschickten eine Erinnerung, 12 Hochschulen zwei Erinnerungen und 3 Hochschulen verschickten drei Erinnerungen an die Befragten.<sup>5</sup> Zum anderen sollte die Teilnahmebereitschaft der Befragten durch Incentives erhöht werden. Hierfür wurden Incentives (Unterhaltungselektronik und Gutscheine) im Gesamtwert von über 2.000€ bereitgestellt.

**[Rücklauf]** Von den Hochschulen wurden 19.330 Exmatrikulierte angeschrieben. Von diesen 19.330 Exmatrikulierten waren 1.051 Personen nachweislich nicht erreichbar. Außerdem wurden 340 Fälle, die nicht der Grundgesamtheit angehören<sup>6</sup>, als stichprobenneutrale Ausfälle vernachlässigt, sodass sich die verbleibende, bereinigte Bruttostichprobe auf 17.939 Exmatrikulierte beläuft. Von diesen gab es insgesamt 4.675 Befragungsteilnahmen (Linkaufrufe). Davon haben 508 Fälle der Einverständniserklärung auf der Startseite nicht oder nicht eindeutig zugestimmt und wurden daher aus den Daten entfernt. Es wurden zudem 563 Fälle gelöscht, die keine gültige Angabe in der für die Befragung besonders zentralen Variablen „absma“ aufweisen. Insgesamt ergibt sich daraus eine Nettostichprobe von 3.604 Fällen. Die Rücklaufquote beträgt somit etwa 20,1 %. Es liegen keine Erkenntnisse zum systematischen Auftreten von Befragungsabbrüchen vor.

Tabelle 1: Brutto-/Nettostichprobengröße und Rücklaufquote

Bruttostichprobe	
Angeschriebene Exmatrikulierte	19.330
davon:	
▪ nicht erreichbar:	1.051
▪ Fälle, die nicht der Grundgesamtheit angehören:	340
Bereinigte Bruttostichprobe insgesamt:	17.939
Nettostichprobe	
Befragungsteilnehmer*innen	4.675
davon:	
▪ keine (eindeutige) Zustimmung zur Einverständniserklärung:	508
▪ keine gültige Angabe in der Variable „absma“:	563
Bereinigte Nettostichprobe insgesamt:	3.604
Rücklaufquote	20,1%

<sup>5</sup> Die Hochschulen verzichteten aufgrund der pandemischen Situation bedingt durch das Corona-Virus ab Mitte März 2020 zum Teil auf zuvor vereinbarte Erinnerungen.

<sup>6</sup> Diese wurden im Nachgang an die Befragung auf Basis der getätigten Angaben identifiziert. Es werden nur die Fälle als Teil der Grundgesamtheit angesehen, für die gilt: absang == 1 ODER (absang < 0 & inlist(absma,2,3))

## 4 Datenaufbereitung

**[Datenübertragung]** Die Angaben aus der Online-Befragung konnten direkt aus der Befragungssoftware als .csv-Datei exportiert und weiterverarbeitet werden.

**[Konsistenzprüfungen]** Wertebereiche und Filterführungen wurden geprüft. Es ergaben sich hieraus jedoch keine Datenmodifikationen. Darauf hinzuweisen ist, dass bei einigen Jahresvariablen, die über eine offene Nennung erfasst wurden, teilweise inkonsistente (oft zweistellige) Werte enthalten sind. Im Datensatzreport wird bei diesen Variablen im Anmerkungsfeld darauf aufmerksam gemacht.

**[Bereinigungen]** Teilweise wurden Bereinigungsmaßnahmen durchgeführt, beispielsweise wurden Anpassungen von Wertelabeln vorgenommen sowie falsch encodete Zeichen in String-Variablen bereinigt.

**[Codierung offener Angaben]** Weiter wurden Nachcodierungen offener Angaben ausgeführt und dabei die folgenden Codes vergeben:

- Hochschulcodes,
- Fächercodes,
- Codes für Geburtsländer und die Staatsangehörigkeit.

Je nach Variable wurden verschiedene Codierlisten und Klassifikationsschlüssel der amtlichen Statistik herangezogen, mit deren Hilfe den offenen Angaben numerische Codes zugeordnet wurden. Die genannten codierten Merkmale wurden im Zuge der Anonymisierung aggregiert (vgl. Kapitel 5).

**[Variablen- und Wertelabelvergabe]** Für Variablen- und Wertelabels wurden Kurzformen der Fragebogenformulierungen gewählt. In den Variablenlabels ist außerdem ein Kürzel für den Zugangsweg enthalten, ab dem die jeweilige Variable nutzbar<sup>7</sup> ist (vgl. Kapitel 1 und 5):

- **c:** Variable ist im CUF und in allen SUF-Varianten (Download-SUF und Remote-Desktop-SUF) nutzbar.
- **r:** Variable ist nicht im CUF und im Download-SUF, aber im Remote-Desktop-SUF nutzbar.
- **a:** Variable ist über keinen Zugangsweg nutzbar. Sie wird aber dokumentiert, da es zugehörige Fragen im Fragebogen gibt.

**[Codierung fehlender Werte]** Bezüglich fehlender Werte wurde einerseits die Systematik der Befragungssoftware Zofar verwendet, andererseits wurden im Zuge der Datenaufbereitung neue Missingwerte vergeben (s. hierzu Tabelle 2). Diese orientieren sich an einer im FDZ-DZHW erstellten übergreifenden Systematik, die über verschiedene Datensätze des DZHW hinweg eine einheitliche Missingcodierung gewährleistet.

---

<sup>7</sup> „Nutzbar“ heißt: die Variable enthält nicht das Missing „anonymisiert“.

Tabelle 2: Verwendete Missingsystematik

Missing-Systematik der Befragungssoftware Zofar	
-9990	nicht beantwortet: Item wurde gesehen, aber nicht beantwortet
-9991	Seite nicht besucht: Aufgrund der Fragebogensteuerung oder eines vorherigen Befragungsabbruches
-9992	nicht gesehen: Item wurde gemäß Fragebogensteuerung nicht angezeigt oder befindet sich auf der Seite des Befragungsabbruches
-9995	Steuerungsvariable fehlt: Variable wurde nicht erhoben (-9990 oder -9991), jedoch für die Fragebogensteuerung verwendet
Missing-Systematik des FDZ-DZHW (angepasst)	
-966	nicht bestimmbar: Eine offene Angabe, die nicht vercodet werden konnte oder für generierte Variablen, die zwar in der Originalvariable einen gültigen Wert aufweisen, für die in der generierten Variable jedoch keine Zuordnung stattfinden kann
-967	anonymisiert: Variable ist in der jeweiligen Zugangsvariante nicht verfügbar
-968	unplausibler Wert
-998	keine Angabe oder divers: Es wurde keine Eingabe getätigt oder divers angegeben (itemspezifisches Missing für das aggregierte Merkmal Geschlecht, vgl. Kapitel 5)

## 5 Anonymisierung

**[Datenschutzrechtlicher Rahmen]** Für personenbezogene Daten<sup>8</sup>, die in freiwilligen Befragungen durch das DZHW erhoben werden, gelten die EU-Datenschutz-Grundverordnung (EU-DSGVO) und das Bundesdatenschutzgesetz in seiner Neufassung vom 30. Juni 2017.<sup>9</sup> Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (ohne Vorliegen einer Einverständniserklärung zur Sekundärnutzung der personenbezogenen Daten) in der Regel derart aufzubereiten, dass „die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden können“ (Art. 4 Abs. 5 DSGVO; s. auch Art. 89 DSGVO sowie Erwägungsgrund 26 DSGVO). Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten derart zu anonymisieren, dass kein Bezug zur Person mehr hergestellt werden kann.

**[Datenzugang, Anonymisierungsgrad und Analysepotential]** Über den *Digital Object Identifier* (DOI) [10.21249/DZHW:stuma2020:1.0.0](https://doi.org/10.21249/DZHW:stuma2020:1.0.0) ist eine Website mit zentralen Informationen zur Studie sowie weiteren Dokumentationsmaterialien erreichbar. Das FDZ des DZHW stellt für StuMa 2020 ein SUF für die wissenschaftliche Sekundärnutzung und ein 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 per Download bereitgestellt wird, wird das SUF über zwei verschiedene Zugangswege angeboten: Download und Remote-Desktop (für weiterführende Informationen vgl. Kapitel 1). Für jeden Zugangsweg wird eine andere SUF-Variante bereitgestellt, die unterschiedlich stark anonymisiert worden ist und entsprechend weniger oder mehr Informationen umfasst. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen in Abhängigkeit von Datenprodukt (SUF/CUF) und Zugangsweg erläutert.

**[Statistische Anonymisierungsmaßnahmen]** Im Rahmen der statistischen Anonymisierungsmaßnahmen wurden zunächst alle Informationen darauf geprüft, ob sich über diese Personen oder Institutionen direkt identifizieren lassen. Im Zuge dessen wurde die Original-Identifikationsnummer entfernt und durch eine neue zufällig vergebene Identifikationsnummer ersetzt. Weitere dieser

---

<sup>8</sup> „Personenbezogene Daten (sind) alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind“ (Art. 4 DSGVO, S. 1).

<sup>9</sup> Die DSGVO gilt grundsätzlich innerhalb der EU und somit ebenfalls für das DZHW. Das BDSG in seiner Neufassung vom 30. Juni 2017 (Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 (Datenschutz-Anpassungs- und Umsetzungsgesetz EU DSAnpUG-EU)) kommt teils zusätzlich 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.

sogenannten *direkten Identifikatoren* wurden in der Befragung nicht erhoben, entsprechend sind sie auch nicht im SUF und CUF der DZHW-Befragung StuMa 2020 enthalten.

Anschließend wurden die *Quasi-Identifikatoren* bestimmt, also Informationen, die in Kombination oder durch die Anspielung externer Informationen geeignet sind, eine Person indirekt zu identifizieren. Um eine eindeutige Zuordnung der Daten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenprodukt bzw. Zugangsweg – aggregiert oder gar nicht freigegeben. Beispielsweise sind für StuMa 2020 zentrale Quasi-Identifikatoren die jeweiligen Hochschulen und Studienfächer. Die Hochschulen werden als Hochschulort und -art aggregiert bereitgestellt. Die Studienfächer werden im Remote-Desktop-SUF aggregiert zu Destatis-Studienbereichen freigegeben. Im Download-SUF und im Download-CUF erfolgt eine noch etwas stärkere Aggregation zu Destatis-Fächergruppen. Im Zuge der Aggregationen neu generierte Variablen sind im Datensatz durch das Suffix „\_g#“ im Variablennamen gekennzeichnet und wurden hinter der jeweiligen Ausgangsvariable positioniert.

Darüber hinaus empfehlen Ebel und Meyermann, offene Angaben zu löschen „selbst wenn die jeweiligen Fragestellungen an sich unproblematisch sind. Denn es besteht die Gefahr, dass Studienteilnehmer/-innen bei eigentlich unbedenklichen Fragen mit offener Antwortmöglichkeit kritische Informationen preisgegeben haben, die zu einer Identifikation führen könnten“ (Ebel & Meyermann, 2015, S. 5). Entsprechend werden bei StuMa 2020 keine offenen Angaben freigegeben.

Zudem ist darauf hinzuweisen, dass bei der Aggregation offener Angaben von Jahreszahlen teils Plausibilisierungsannahmen getroffen wurden. Hierzu wurden Kurzformen der angegebenen Jahre zu den jeweiligen Jahren aggregiert (bspw. wird „12“ dem Jahr 2012 zugeordnet). Den Datennutzer\*innen bleibt es so offen, die vom FDZ-DZHW plausibilisierte Variable zu nutzen oder die unplausibilisierte Variante. Plausibilisierte Variablen sind im Datensatz durch das Suffix „\_p#“ im Variablennamen gekennzeichnet.

Zuletzt wurde geprüft, ob in den Daten sensible Informationen, z. B. zur Gesundheit, sexuellen Orientierung und zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht notwendig zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (vgl. Koberg, 2016a, S. 694) und sind daher besonders schützenswert (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). Einzelinformationen zur Gesundheit werden im Rahmen von StuMa 2020 nicht freigegeben.

Im CUF wurden dieselben statistischen Anonymisierungsmaßnahmen auf Variablenebene wie im Download-SUF vorgenommen, jedoch wurde eine per Zufallsauswahl gewonnene Substichprobe der Daten (50 % der Befragten) gezogen. Die nachfolgende Tabelle 3 stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen je nach Datenform bzw. Zugangsweg dar. Genauere Informationen zur Aggregation der einzelnen Variablen sind im Datensatzreport zu finden

Tabelle 3: Maßnahmen der statistischen Anonymisierung der Daten von StuMa 2020

Merkmal	Remote-Desktop-SUF	Download-SUF/ Download-CUF (Substichprobe)
Direkte Identifikatoren	Löschung/ Vergabe einer zufälligen ID	Löschung/ Vergabe einer zufälligen ID
Monatsangaben (offene Angabe)	Aggregation zu den Kategorien: - „Januar/Februar“ - „März/April“ - „Mai/Juni“ - „Juli/August“ - „September/Okttober“ - „November/Dezember“	Löschung
Ausnahmen: Beginn- und Endmonat des dritten vorherigen Studiums	Löschung	Löschung
Geburtsmonat	Freigabe	Löschung
Erste Immatrikulation Master (Jahr) (offene Angabe)	Freigabe	
	Aggregation zu den Kategorien: - „1988-2008“ - „2009-2012“ 2013-2019 werden einzeln ausgewiesen	Aggregation zu den Kategorien: - „1988-2008“ - „2009-2012“ 2013-2019 werden einzeln ausgewiesen
Hochschule (offene Angabe)	Vercodung und Aggregation zum Hochschulort (Bundesländergruppen) <sup>a</sup> : - „Baden-Württemberg“ - „Bayern“ - „Hessen“ - „Nordrhein-Westfalen“ - „Berlin, Brandenburg, Mecklenburg-Vorpommern“ - „Bremen, Niedersachsen“ - „Hamburg, Schleswig-Holstein“ - „Rheinland-Pfalz, Saarland“ - „Sachsen, Sachsen-Anhalt, Thüringen“	
	Vercodung und Aggregation zum Hochschulort (alte/neue Bundesländer) <sup>a</sup>	Vercodung und Aggregation zum Hochschulort (alte/neue Bundesländer) <sup>a</sup>
	Vercodung und Aggregation zur Hochschulart <sup>a</sup> : - Universität - Fachhochschule (FH)	Vercodung und Aggregation zur Hochschulart <sup>a</sup> : - Universität - Fachhochschule (FH)
Studienfach (offene Angabe)	Vercodung und Aggregation zu Studienbereichen <sup>b</sup>	
	Vercodung und Aggregation zur Fächergruppe <sup>b</sup>	Vercodung und Aggregation zur Fächergruppe <sup>b</sup>
Letzte Immatrikulation Master (Jahr) (offene Angabe)	Freigabe	
	Aggregation zu den Kategorien: - „2000-2010“ - „2011-2013“ 2014-2019 werden einzeln ausgewiesen	Aggregation zu den Kategorien: - „2000-2010“ - „2011-2013“ 2014-2019 werden einzeln ausgewiesen
Note (offene Angabe)	Freigabe	
	Rundung zu vollen Noten	Rundung zu vollen Noten
Grund für das Beenden des Studiums ohne Abschluss: Krankheit/psychische Probleme	Löschung	Löschung

Höhe der finanziellen Quellen (offene Angabe)	Freigabe  Aggregation in 100er-Schritten sowie Top-Codierung ab 5000 Euro: - „bis 100 Euro“ - „101-200 Euro“ - „201-300 Euro“ - (usw.) - „5000 Euro und mehr“	Aggregation in 100er-Schritten sowie Top-Codierung ab 5000 Euro: - „bis 100 Euro“ - „101-200 Euro“ - „201-300 Euro“ - (usw.) - „5000 Euro und mehr“
Beginn des zum Master berechtigenden Studiums (Jahr) (offene Angabe)	Freigabe  Aggregation zu den Kategorien: - „1980-2004“ - „2005-2008“ - „2018/19“ 2009-2017 werden einzeln ausgewiesen und unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.  Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).	Aggregation zu den Kategorien: - „1980-2004“ - „2005-2008“ - „2018/19“ 2009-2017 werden einzeln ausgewiesen und unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.  Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).
Ende des zum Master berechtigenden Studiums (Jahr) (offene Angabe)	Freigabe  Aggregation zu den Kategorien: - „1983-2008“ - „2009-2013“ 2014-2019 werden einzeln ausgewiesen und unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.  Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).	Aggregation zu den Kategorien: - „1983-2008“ - „2009-2013“ 2014-2019 werden einzeln ausgewiesen und unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.  Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).
Beginn des zweiten vorherigen Studiums (Jahr) (offene Angabe)	Freigabe  Aggregation zu den Kategorien: - „1980-2005“ - „2006/07“ - „2008/09“ - „2010/11“ - „2012/13“ - „2014/15“ - „2016-2019“ Unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.  Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).	Aggregation zu den Kategorien: - „1980-2005“ - „2006/07“ - „2008/09“ - „2010/11“ - „2012/13“ - „2014/15“ - „2016-2019“ Unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.  Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).
Ende des zweiten vorherigen Studiums (Jahr) (offene Angabe)	Freigabe  Aggregation zu den Kategorien: - „1984-2007“ - „2008/09“ - „2010/11“ - „2012/13“ - „2014/15“ - „2016/17“ - „2018/19“ Unplausible Werte werden zur	Aggregation zu den Kategorien: - „1984-2007“ - „2008/09“ - „2010/11“ - „2012/13“ - „2014/15“ - „2016/17“ - „2018/19“ Unplausible Werte werden zur

	Kategorie „unplausibler Wert“ aggregiert.	Kategorie „unplausibler Wert“ aggregiert.
	Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).	Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).
Beginn des dritten vorherigen Studiums (Jahr) (offene Angabe)	Freigabe  Aggregation zu den Kategorien: - „1996-2005“ - „2006-2010“ - „2011-2017“ Unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.	Aggregation zu den Kategorien: - „1996-2005“ - „2006-2010“ - „2011-2017“ Unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.
	Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).	Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).
Ende des dritten vorherigen Studiums (Jahr) (offene Angabe)	Freigabe  Aggregation zu den Kategorien: - „1997-2007“ - „2008-2014“ - „2015-2019“ Unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.	Aggregation zu den Kategorien: - „1997-2007“ - „2008-2014“ - „2015-2019“ Unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.
	Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).	Plausibilisierungsannahmen: Kurzformen von Jahren werden zu den jeweiligen Jahren aggregiert (z.B. 12 wird dem Jahr 2012 zugeordnet).
Erwerb der Hochschulzugangsberechtigung (Jahr) (offene Angabe)	Freigabe  Aggregation zu den Kategorien: - „1978-1999“ - „2000-2007“ - „2016-2019“ 2008-2015 werden einzeln ausgewiesen	Aggregation zu den Kategorien: - „1978-1999“ - „2000-2007“ - „2016-2019“ 2008-2015 werden einzeln ausgewiesen
Bundesland des Erwerbs der Hochschulzugangsberechtigung	Aggregation zu Bundesländergruppen: - „Baden-Württemberg“ - „Bayern“ - „Hessen“ - „Nordrhein-Westfalen“ - „Berlin, Brandenburg, Mecklenburg-Vorpommern“ - „Bremen, Niedersachsen“ - „Hamburg, Schleswig-Holstein“ - „Rheinland-Pfalz, Saarland“ - „Sachsen, Sachsen-Anhalt, Thüringen“	
	Aggregation zu alte/neue Bundesländer	Aggregation zu alte/neue Bundesländer
Staat des Erwerbs der Hochschulzugangsberechtigung (offene Angabe)	Vercodung und Aggregation nach der NEPS-Länderzuordnung (ohne die Einzelausweisung der Länder: Italien, Polen, Rumänien und Türkei) <sup>c</sup>	
	Vercodung und Aggregation zu Weltregionen <sup>c</sup> :	Vercodung und Aggregation zu Weltregionen <sup>c</sup> :

	<ul style="list-style-type: none"> <li>- „Afrika“</li> <li>- „Amerika“</li> <li>- „Asien“</li> <li>- „Australien/Ozeanien“</li> <li>- „EU“</li> <li>- „Europa außerhalb der EU“</li> </ul>	<ul style="list-style-type: none"> <li>- „Afrika“</li> <li>- „Amerika“</li> <li>- „Asien“</li> <li>- „Australien/Ozeanien“</li> <li>- „EU“</li> <li>- „Europa außerhalb der EU“</li> </ul>
Geburtsjahr (offene Angabe)	Freigabe  Aggregation zu den Kategorien: <ul style="list-style-type: none"> <li>- „1959-1977“</li> <li>- „1978/79“</li> <li>- „1980/81“</li> <li>- „1982/83“</li> <li>- „1984/85“</li> <li>- „1986/87“</li> <li>- „1996-2005“</li> </ul> 1988-1995 werden einzeln ausgewiesen und unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.	Aggregation zu den Kategorien: <ul style="list-style-type: none"> <li>- „1959-1977“</li> <li>- „1978/79“</li> <li>- „1980/81“</li> <li>- „1982/83“</li> <li>- „1984/85“</li> <li>- „1986/87“</li> <li>- „1996-2005“</li> </ul> 1988-1995 werden einzeln ausgewiesen und unplausible Werte werden zur Kategorie „unplausibler Wert“ aggregiert.
Geschlecht	Freigabe  Aggregation der Kategorien „divers“ und „nicht beantwortet“ zu „keine Angabe oder divers“	Aggregation der Kategorie „divers“ und „nicht beantwortet“ zu „keine Angabe oder divers“
Kinderanzahl (offene Angabe)	Top-Codierung: „4 und mehr Kinder“, ansonsten einzeln ausgewiesen.	Top-Codierung: „4 und mehr Kinder“, ansonsten einzeln ausgewiesen.
Geburtsland und Staatsangehörigkeit der Befragten und der Eltern der Befragten (offene Angabe)	Vercodung und Aggregation nach der NEPS-Länderzuordnung (bei Befragten: ohne die Einzelausweisung der Länder: Italien, Polen, Rumänien und Türkei) <sup>c</sup>  Vercodung und Aggregation zu Weltregionen <sup>c</sup> : <ul style="list-style-type: none"> <li>- „Afrika“</li> <li>- „Amerika“</li> <li>- „Asien“</li> <li>- „Australien/Ozeanien“</li> <li>- „EU“</li> <li>- „Europa außerhalb der EU“</li> </ul>	Vercodung und Aggregation zu Weltregionen <sup>c</sup> : <ul style="list-style-type: none"> <li>- „Afrika“</li> <li>- „Amerika“</li> <li>- „Asien“</li> <li>- „Australien/Ozeanien“</li> <li>- „EU“</li> <li>- „Europa außerhalb der EU“</li> </ul>
Migrationsjahr (offene Angabe)	Löschung aufgrund uneinheitlicher, oft unplausibler Angaben	Löschung aufgrund uneinheitlicher, oft unplausibler Angaben
Sonstige offene Nennungen	Löschung	Löschung

<sup>a</sup> gemäß Destatis-Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik, Wintersemester 2019/2020 und Sommersemester 2020

<sup>b</sup> gemäß Destatis-Fächersystematik, Wintersemester 2019/2020 und Sommersemester 2020

<sup>c</sup> vgl. cl-dzhw-47 ([https://metadata.fdz.dzhw.eu/public/files/instruments/ins-stuma2020-ins1\\$/attachments/cl-dzhw-47.xls](https://metadata.fdz.dzhw.eu/public/files/instruments/ins-stuma2020-ins1$/attachments/cl-dzhw-47.xls)) (Die NEPS-Länderzuordnung basiert auf der folgenden Veröffentlichung mit Anpassungen: [https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/6-0-0/SC5\\_6-0-0\\_Anonymisation.pdf](https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/6-0-0/SC5_6-0-0_Anonymisation.pdf), S. 33-34)

## 6 Literatur

- Ebel, T. & Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten. Forschungsdaten Bildung informiert. Bd.3. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung.
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). Datenschutz am Forschungszentrum (FDZ-Methodenreport Nr. 6). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- 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
- 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.
- Schupp, J., & Gerlitz, J.-Y. (2014). Big Five Inventory-SOEP (BFI-S). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). <https://doi.org/10.6102/zis54>