

Daten- und Methodenbericht
Juli 2023

Henrike Schmidtchen | Thorsten Euler | Tina Oestreich

Absolventen- befragung 2017

Daten- und Methodenbericht zur Erhebung der
Absolvent*innenkohorte 2017

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/>



Projektmitarbeiter*innen

Henrike Schmidtchen

Thorsten Euler

Tina Oestreich

Impressum

Herausgeber

Deutsches Zentrum für Hochschul- und
Wissenschaftsforschung GmbH (DZHW)

Lange Laube 12 | 30159 Hannover | www.dzhw.eu

Postfach 2920 | 30029 Hannover

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

Geschäftsführung

Prof. Dr. Monika Jungbauer-Gans

Vorsitzender des Aufsichtsrats

Ministerialdirigent Peter Greisler

Registergericht

Amtsgericht Hannover | B 210251

Umsatzsteuer-Identifikationsnummer:

DE291239300

Juli 2023

Inhalt

Tabellen-/Abbildungsverzeichnis	I
I. Einleitung	2
1 Inhalt und Anlage der Studien	4
2 Erhebungsinstrument	7
3 Grundgesamtheit, Stichprobenumfang und Rücklauf	9
4 Datenaufbereitung	11
4.1 Datenübertragung.....	11
4.2 Codierung offener Angaben.....	11
4.3 Datenbereinigung	12
4.4 Referenzstudium und Studienepisoden.....	12
4.5 Generierung von Variablen.....	13
4.6 Erstellung des Datensatz.....	14
4.7 Vergabe von Variablenamen, Variablenlabels und Wertelabels	14
4.8 Codierung fehlender Werte	15
5 Anonymisierung	16
6 Literatur	21
Anhang 1	22

Tabellen-/Abbildungsverzeichnis

Abbildung 1:	Kohorten-Panel-Design der DZHW-Absolventenstudienreihe 1989-2021.....	4
Tabelle 1:	Anlage der Hauptuntersuchungen der DZHW-Absolventenstudienreihe von 1989 bis 2021	5
Tabelle 2:	Vercodete Merkmale und Codierlisten in der Absolventenbefragung 2017	12
Tabelle 3:	Systematik des Primärforschungsprojekts für fehlende Werte	15
Abbildung 3:	Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten der Absolventenbefragung 2017.....	17
Tabelle 4:	Maßnahmen der statistischen Anonymisierung der Daten der Absolventenbefragung 2017 nach Zugangsweg	18

I. Einleitung

Seit 1989 wird alle vier Jahre ein Absolvent*innenjahrgang ausgewählt und nach einem, vier und zum Teil auch nach zehn Jahren befragt.¹ Sie wurden bis einschließlich der Absolvent*innenkohorte 2013 durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW) durchgeführt, vom Bundesministerium für Bildung und Forschung (BMBF) gefördert und dienen – in Ergänzung zur amtlichen Hochschulstatistik – dem nationalen Bildungsmonitoring.

Mit der Absolventenbefragung 2017 wurde erstmals ein bundesweites Projekt der deutschen Absolventenforschung von einem Verbund überregional tätiger Forschungsinstitute durchgeführt: International Centre for Higher Education Research (INCHER), Institut für angewandte Statistik (ISTAT) und Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) (Fabian, Flöther & Reifenberg, 2021). Der Grund hierfür waren insbesondere Erfahrungen der DZHW-Absolventenstudie des Jahrgangs 2013, die vom systematischen Ausfall einzelner Bundesländer in der Stichprobe u.a. aufgrund der Beteiligung dortiger Hochschulen an der KOAB-Erhebung desselben Jahrgangs betroffen war (siehe hierzu: Hoffstätter, Niebuhr & Vietgen, 2021). Ab der Kohorte 2021 wird die Reihe der DZHW-Absolventenstudien wieder in alleiniger Verantwortung des DZHW im Rahmen des Student Life Cycle Panel (SLC) fortgesetzt.

Innerhalb des Verbundprojekts der Absolventenbefragung 2017 war das INCHER für die Informations- und Öffentlichkeitsarbeit sowie die Dissemination der Ergebnisse zuständig. Dem ISTAT oblag die Administration des Feldzugangs inklusive Akquise der Befragungshochschulen, die Programmierung des Fragebogens, die Durchführung der Befragung und die prinzipielle Datenaufbereitung. Das DZHW war für das Stichprobenkonzept, das Berichtswesen und die Projektkoordination verantwortlich. Detailplanung und Datenanalyse erfolgten im Verbund. Erste Ergebnisse des Projekts wurden von den Verbundpartnern unter dem Titel „Generation Hochschulabschluss: neue und alte Differenzierungen“ (Fabian, Flöther & Reifenberg, 2021) publiziert.

Zum Kuratieren der Daten wurden diese abschließend an das Forschungsdatenzentrum für Hochschul- und Wissenschaftsforschung am DZHW (FDZ-DZHW) übergeben. Dort werden die Daten vorhergehender Absolventenkohorten ab der Kohorte 1989 sowie der vorliegenden Absolventenbefragung 2017 nachträglich zum Zweck der Datennachnutzung aufbereitet und dokumentiert. Grundlage der im FDZ-DZHW vorgenommenen Datenaufbereitung spezifisch der Absolventenbefragung 2017 stellte ein vom ISTAT voreditierter Datensatz dar. Die durch das FDZ-DZHW weiter aufbereiteten Daten werden über verschiedene Zugangswege als *Scientific Use Files* (SUF) für die wissenschaftliche Sekundärnutzung und 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.

Im Vergleich zu den DZHW-Absolventenstudien vorheriger Jahre gelten für die Absolventenbefragung 2017 besondere Einschränkungen, welche insbesondere in Kapitel 1 im Abschnitt „Analysepotential“ und Kapitel 4 nachvollzogen werden können. Zentral zu nennen sind starke Einschränkungen der Qualität und Nutzbarkeit der Tätigkeits-, Studien- und Berufsverläufe. Im Gegensatz zu den früheren durch

¹ Informationen zur Absolventenstudienreihe können über die Website des Projektes (www.dzhw.eu/absolventen) abgerufen werden.

das DZHW erhobenen Absolventenpaneldaten können in der vorliegenden Absolventenbefragung 2017 keine Episodendaten zur Verfügung gestellt werden. Die Absolventenbefragung 2017 mit der Datenerhebung durch ISTAT ist zudem keine Längsschnittbefragung im Paneldesign, da keine eindeutige Panel-ID der Befragten generiert wurde. Insgesamt sollte die Absolventenbefragung 2017 nicht uneingeschränkt als weitere Kohorte der Absolventenstudien des DZHW betrachtet werden. Sie ist eher eine im Verbundprojekt durchgeführte Absolventenstudie in deren Kontext, die verglichen zu den Absolventenstudien der früheren Kohorten die genannten Limitationen aufweist. Zugleich sind neben diesen kritischen auch positive Aspekte der zur Sekundärnutzung zur Verfügung stehenden Daten zu nennen. Es wurde ein zwischen den Verbundpartnern abgestimmtes Instrument entwickelt, das die Bearbeitung von Fragen des nationalen Bildungsmonitorings, der Hochschulsteuerung und Hochschulforschung erlaubt (Fabian, Flöther & Reifenberg, 2021). Bei der Absolventenbefragung 2017 stehen der Übergang in den Arbeitsmarkt, in weitere akademische Qualifikationen und eine erste Arbeitsmarktplatzierung im Fokus. Adressierte Themen sind beispielsweise die vorhochschulische Bildung, der Studienverlauf und abgeschlossenes Studium, weitere akademische Qualifikationen und der Berufsstart sowie die berufliche Situation der Absolvent*innen, Kompetenzstände zum Studienabschluss, Kompetenzanforderungen der gegenwärtig Erwerbstätigen sowie die Berufs- und Lebensziele der Hochschulabsolvent*innen.

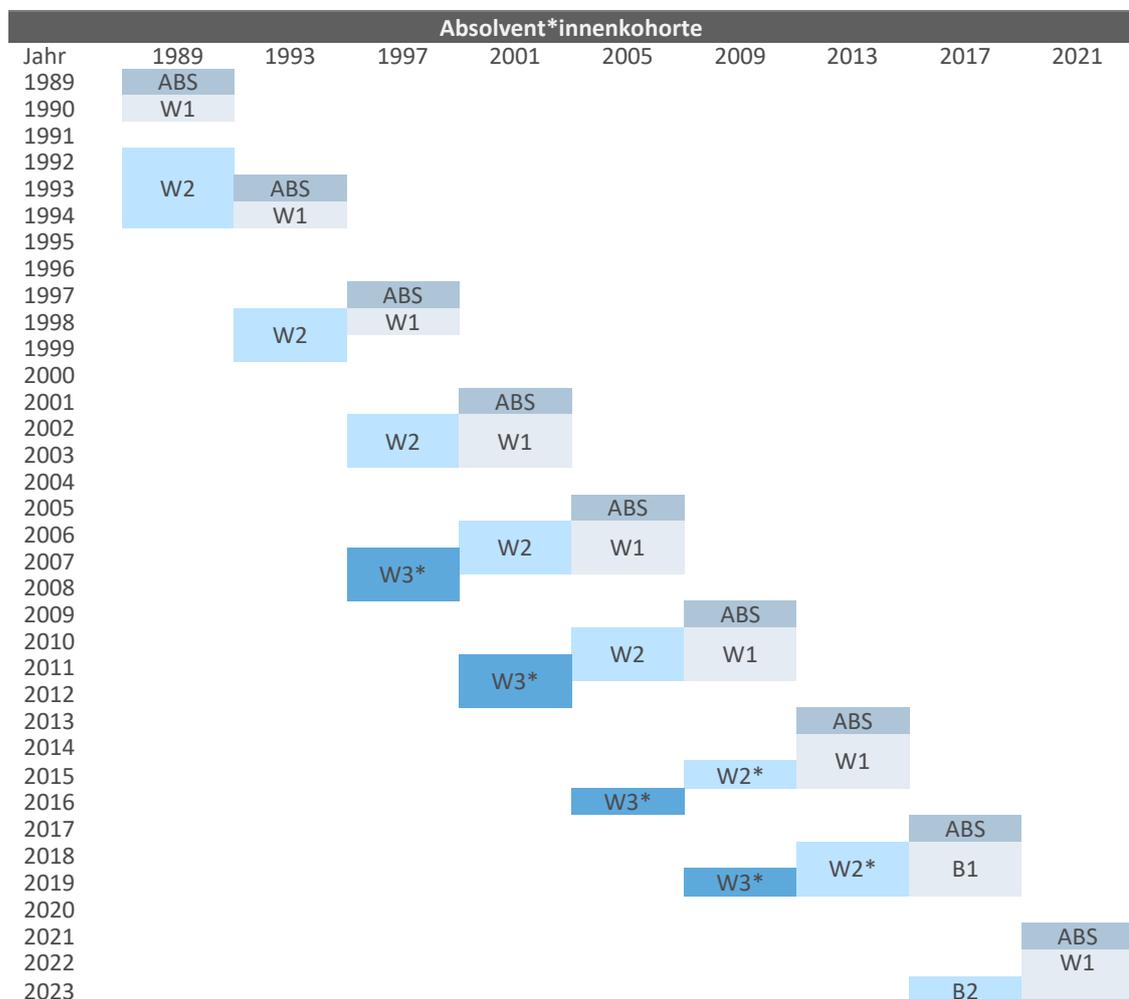
Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation zur ersten Befragung der Absolvent*innenkohorte 2017 ([doi: 10.21249/DZHW:gra2017:1.0.0](https://doi.org/10.21249/DZHW:gra2017:1.0.0)). Weitere Dokumentationsmaterialien zur Studie (Datensatzreports, Fragebögen, Filterführungsdiagramme etc.) können frei im Metadatensuchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden.

Kapitel 1 stellt Inhalt und Anlage der DZHW-Absolventenstudienreihe im Allgemeinen und der Absolventenbefragung 2017 im Speziellen vor. In Kapitel 2 bis 4 wird der Erhebungsprozess (vom Erhebungsinstrument über die Stichprobenziehung bis hin zur Datenaufbereitung) beschrieben. Im Kapitel 5 erfolgt die Beschreibung der vorgenommenen Anonymisierungsmaßnahmen.

1 Inhalt und Anlage der Studien

[Studienreihe] Die Absolventenbefragung 2017 erfolgt im Kontext der DZHW-Absolventenstudienreihe, in der anhand von standardisierten Befragungen Informationen zu Studium, Berufseintritt, Berufsverlauf und Weiterqualifizierung von Hochschulabsolvent*innen erfasst werden. Das erste Absolventenpanel wurde 1989² durchgeführt, seitdem wird jeder vierte Absolvent*innenjahrgang (Kohorte) untersucht.

Abbildung 1: Kohorten-Panel-Design der DZHW-Absolventenstudienreihe 1989-2021



*Hauptbefragung + Vertiefungsbefragungen

**Abkürzungen: ABS = Abschluss, W1 = Welle 1, W2 = Welle 2, W3 = Welle 3, B1 = 1. Befragung, B2 = 2. Befragung

Die Grundgesamtheit einer Kohorte sind Hochschulabsolvent*innen, die im Winter- oder Sommersemester des betreffenden Prüfungsjahrs ihren Studienabschluss an einer Hochschule in der

² Schon seit 1974 werden Hochschulabsolvent*innen – neben den Studienabbrecher*innen und Hochschulwechsler*innen – im Rahmen der Exmatrikuliertenbefragung des DZHW befragt. Diese Untersuchungsreihe wird seit Anfang der 2000er Jahre unter dem Namen „Studienabbruch – Umfang und Motive“ durchgeführt.

Bundesrepublik Deutschland erworben haben.³ Für jede Absolvent*innenkohorte werden mehrere Befragungswellen durchgeführt, wobei jede Welle in unterschiedlichem zeitlichem Abstand zum Studienabschluss stattfindet. Es handelt sich somit um ein kombiniertes Kohorten-Panel-Design (vgl. Abbildung 1).

Die Befragungen der Absolvent*innenkohorten von 1989 und 1993 umfassten zwei Wellen; seit 1997 wird eine zusätzliche dritte Befragungswelle durchgeführt. Die erste Welle einer Kohorte findet im Mittel ein Jahr nach dem jeweiligen Studienabschluss statt. Die zweite Befragungswelle folgt etwa fünf Jahre nach dem Studienabschluss. Etwa zehn Jahre nach dem Abschluss schließt sich die dritte Befragungswelle an. Teilweise setzen sich die zweiten bzw. dritten Wellen aus einer Hauptbefragung und gesonderten Vertiefungsbefragungen zu spezifischen Themen zusammen. Ab der Kohorte 2021 werden im Rahmen des SLC-Projekts zusätzlich zwischen den Wellen Erhebungen durchgeführt, so dass dann insgesamt fünf Befragungen je Kohorte anfallen.

Die verschiedenen Erhebungen wurden vorwiegend als schriftlich-postalische Paper-and-Pencil-Befragung (Paper and Pencil Interview; PAPI) durchgeführt, in den letzten Jahren aber zunehmend auch als Online-Befragung (Computer Assisted Web Interview; CAWI) (vgl. Tabelle 1).

Tabelle 1: Anlage der Hauptuntersuchungen der DZHW-Absolventenstudienreihe von 1989 bis 2021

Welle	Befragungszeitpunkt	Thematischer Fokus	Absolvent*innenkohorte ^c						
			1989 u. 1993	1997 u. 2001	2005	2009	2013	2017	2021
1	ca. 1 Jahr nach Studienabschluss	Studienverlauf und -erfahrung, akademische Weiterqualifizierung, Übergang in den Beruf	Paper & Pencil	Paper & Pencil	Paper & Pencil	Paper & Pencil	Paper & Pencil	Online	Online
2	ca. 5 Jahre nach Studienabschluss	Aktuelle Tätigkeit, Erwerbstätigkeit, akademische und berufliche Weiterbildung	Paper & Pencil	Paper & Pencil	Paper & Pencil	Online ^b	Online ^b	Online	Online ^d (geplant)
3	ca. 10 Jahre nach Studienabschluss	Aktuelle Tätigkeit, Erwerbstätigkeit, akademische und berufliche Weiterbildung, familiäre Situation	---	Paper & Pencil ^a	Online ^b	Online ^b	Online (geplant)	Online (geplant)	Online ^d (geplant)

- a Die Hauptbefragung wurde als Paper & Pencil- und die Vertiefungsbefragungen als Online-Befragung durchgeführt.
- b Sowohl die Hauptbefragung als auch die Vertiefungsbefragungen wurden als Online-Befragung durchgeführt
- c Die Kohorten 1989 bis 2013 wurden in alleiniger Verantwortung des DZHW bzw. dessen Vorgängerorganisation HIS eG durchgeführt. Die Kohorte 2017 wurde im Rahmen eines Institutionenverbunds realisiert. Ab Kohorte 2021 wird das DZHW wieder in alleiniger Verantwortung befragen und die Absolventenstudien in das Student Life Cycle Panel (SLC) integriert.
- d Zusätzlich ist im SLC jeweils eine weitere Befragung zwischen den Wellen 1 und 2 bzw. 2 und 3 vorgesehen, so dass eine zwei-jährige Taktung mit insgesamt fünf Befragungen erreicht wird.

Die Erhebungsinstrumente aller Kohorten enthalten Fragen zum Studium, zum Übergang in den Beruf, zur akademischen und beruflichen Weiterbildung, zur Erwerbstätigkeit und zu soziodemographischen und bildungsbiographischen Merkmalen. Der thematische Fokus einer Befragungswelle orientiert sich an der jeweils typischen Bildungs-, Berufs- und Lebensphase der Befragten zum Befragungszeitpunkt.

³ Für die Kohorte 1989 wurden ausschließlich Absolvent*innen aus den Bundesländern der damaligen Bundesrepublik Deutschland ausgewählt.

[Analysepotential] Aufgrund eines neu konzipierten Erhebungsinstruments im Rahmen des Verbundprojekts der Absolventenbefragung 2017 ergibt sich ein geringeres Analysepotential. Auch wenn in dieser Kohorte teilweise ein identischer Kern an Informationen erhoben wurde, können Zeitreihen- bzw. Kohortenvergleiche zur Abbildung langfristiger Trends der Hochschulbildung und Arbeitsmarktentwicklung nur bedingt in den Blick genommen werden. Weiterhin sollte beachtet werden, dass aufgrund der Nicht-Erhebung einer Panelvariable in der Absolvent*innenkohorte 2017 keine intra-individuellen Veränderungen zwischen den Wellen (z.B. kausale Panelanalysen) betrachtet werden können, Nettoveränderungen bezogen auf die gesamte Stichprobe können jedoch dargestellt werden. Die Absolventenbefragung 2017 sollte entsprechend nicht als eine weitere Kohorte in den DZHW-Absolventenstudienreihen, sondern als alleinstehende Erhebung im Kontext der DZHW-Absolventenstudien betrachtet werden. Dennoch werden in Abhängigkeit von aktuellen Entwicklungen und Forschungsinteressen in einzelnen Kohorten bestimmte Aspekte vertiefend oder ergänzend abgefragt – so auch bei der Absolventenbefragung 2017.

[Einordnung ins Forschungsfeld] Das Stichproben- und Erhebungsdesign sowie die damit verbundenen Analysemöglichkeiten unterscheiden die DZHW-Absolventenstudienreihe von anderen in Deutschland durchgeführten Absolventenstudien. So ist beispielsweise das Bayerische Absolventenpanel (BAP) des Bayerischen Staatsinstituts für Hochschulforschung und Hochschulplanung (IHF) auf Absolvent*innen bayerischer Hochschulen begrenzt⁴. Das Kooperationsprojekt Absolventenstudie (KOAB) des International Centre for Higher Education Research (INCHER) bzw. des Instituts für angewandte Statistik (ISTAT) befragt Absolvent*innen seiner Kooperationshochschulen und ermöglicht individuelle Analysen auf Hochschul- und Studiengangsebene, die zur Evaluation und Weiterentwicklung genutzt werden können, stellt jedoch kein bundesweit repräsentatives Sample dar, da Selbstselektion der Hochschulen in die Befragung vorliegt.⁵

[Spezifika der Absolventenbefragung 2017] Die Absolventenbefragung 2017 ist das erste Verbundprojekt im Kontext der DZHW-Absolventenstudienreihe. Die Verbundpartner DZHW, INCHER, ISTAT haben das Ziel, Interessen unterschiedlicher Nutzer*innengruppen möglichst umfassend zu berücksichtigen. Hierzu soll ein gemeinsam abgestimmtes und neukonzipiertes Instrument ermöglichen, Fragen sowohl des nationalen Bildungsmonitorings als auch der Hochschulsteuerung und Hochschulforschung zu berücksichtigen (Fabian, Flöther & Reifenberg, 2021).

Thematisch stehen bei der Absolventenbefragung 2017 der Übergang in den Arbeitsmarkt (oder in weitere akademische Qualifikationen) sowie eine erste Arbeitsmarktplatzierung im Mittelpunkt. Es werden außerdem verschiedenste, weitere Forschungsthemen in der Erhebung adressiert:

- vorhochschulische Bildung,
- Studienverlauf und abgeschlossenes Studium,
- weitere akademische Qualifikationen,
- Berufsstart und die berufliche Situation der Absolvent*innen.

Darüber hinaus wurden die Kompetenzstände zum Studienabschluss, die Kompetenzanforderungen der gegenwärtig Erwerbstätigen sowie die Berufs- und Lebensziele der Hochschulabsolvent*innen erfragt.

⁴ vgl. dazu <http://www.bap.ihf.bayern.de>

⁵ vgl. dazu <http://koab.uni-kassel.de>

2 Erhebungsinstrument

In der Absolventenbefragung 2017 wurden Kenntnisse aus den DZHW-Absolventenstudien mit dem Kooperationsprojekt Absolventenstudie (KOAB) in einem neu konzipierten und standardisierten Onlinefragebogen vereint. Das KOAB wird vom ISTAT geleitet, welches auch die technische Infrastruktur für die Befragung bereitgestellt und diese durchgeführt hat⁶. Insbesondere wurden neben einer allgemeinen Version des Befragungsinstruments verschiedenen Fragebogenvarianten, welche im KOAB Verwendung finden, genutzt. Bei den Varianten handelt es sich um Modifikationen der eingesetzten Fragebögen, welche die Situation bestimmter Absolvent*innengruppen spezifisch berücksichtigen. Diese Varianten sollen im Rahmen der Differenzierung der Hochschullandschaft hinsichtlich der Abschlussarten, Studienformate und Studieninhalte behilflich sein. Die folgenden Fragebogenvarianten wurden daher mit leichten Anpassungen in die Absolventenbefragung 2017 übernommen: Lehramt, Medizin, Jura, Promotion, Kunst (und Musik), sowie nicht-traditionelle Studienformate. Die Varianten werden jedoch nicht vollständig im SUF/CUF mitberücksichtigt. Lediglich Variablen der allgemeinen Fragebogenversion, die ebenfalls in den Varianten erhoben wurden, sind im SUF/CUF enthalten. Dies sollte in weiteren Analysen berücksichtigt werden. Das Instrument wurde den Befragten in deutscher und englischer Sprache angeboten.

Die teilnehmenden Hochschulen stellten im Vorfeld der Befragung dem ISTAT Daten wie Abschlussart und Studienfächer je Absolvent*in (pseudonymisiert) zur Verfügung. Diese wurden zur Zuweisung der Befragten zur jeweils spezifischen Fragebogenvarianten, als eingeblendete individualisierte Informationen (Preloads) zu Beginn des Fragebogens und teilweise für Angaben in den Variablen zum Referenzstudium verwendet.

Folgende Themenbereiche wurden in der allgemeinen Variante der Absolventenbefragung 2017 abgefragt:

- Vor dem Studium
- Studium (Referenzstudium)
- Übergang in weitere akademische Qualifikation: Master-Studium
- Übergang in weitere akademische Qualifikation: Promotion
- Weitere Studienepisoden
- Auslandserfahrung
- Übergang Studium – Beruf
- Berufliche Situation
- Soziodemographie
- Persönlichkeit/Ziele.

Gleichzeitig hatten einzelne Hochschulen durch die Kompatibilität des Kernfragebogens zum Befragungsdesign von KOAB die Möglichkeit, den Kernfragebogen um eigene, individuelle Fragen zu ergänzen.

⁶ Hierunter fällt auch die Adresserfassung für Wiederholungsbefragungen.

[Sekundär genutzte Erhebungsinstrumente] In der Absolventenbefragung 2017 kam die Big-5-Kurzskala BFI-10 zum Einsatz (Variablen: k1bigf*). Siehe hierzu:

- Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less. A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212. doi:10.1016/j.jrp.2006.02.001
- Rammstedt, B. (2007). The 10-Item Big Five Inventory. Norm values and investigation of socio-demographic effects based on a German population representative sample. *European Journal of Psychological Assessment*, 23(3), 193–201. doi:10.1027/1015-5759.23.3.193
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2013). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit. 10 Item Big Five Inventory (BFI-10). *methoden, daten, analysen*, 7(2), 233–249. doi:10.12758/mda.2013.013
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2014). Big Five Inventory (BFI-10). Zusammenstellung sozialwissenschaftlicher Items und Skalen. doi:10.6102/zis76

[Feldphase] Der Erhebungszeitraum der ersten Erhebung erstreckte sich zwischen Oktober 2018 und April 2019. Die Kontaktaufnahme und der Versand der Befragungsunterlagen erfolgte aus Datenschutzgründen im Adressmittlungsverfahren über die beteiligten Hochschulen, welche die Absolvent*innen je nach Bedarf postalisch oder per E-Mail einluden. In der Regel wurden außerdem zwei weitere Erinnerungen an die Absolvent*innen gesendet, wenn diese die Befragung noch nicht begonnen oder unvollständig ausgefüllt hatten.

3 Grundgesamtheit, Stichprobenumfang und Rücklauf

[Grundgesamtheit] Die Grundgesamtheit der Absolventenbefragung 2017 umfasst alle Absolvent*innen, die im Prüfungsjahr 2017 (Wintersemester 2016/2017 und Sommersemester 2017) ihren ersten berufsqualifizierenden Hochschulabschluss oder einen Masterabschluss an einer staatlich anerkannten Hochschule in Deutschland erworben haben (mit Ausnahme der Verwaltungsfachhochschulen, Berufsakademien, die keine akademischen Abschlüsse anbieten, und Hochschulen der Bundeswehr). Darunter fallen alle Bachelor-, Master-, Staatsexamens- und künstlerischen Abschlüsse. Ausgenommen von der Grundgesamtheit sind Personen, die ihr Studium mit einem Diplom- oder Magisterabschluss beendet haben.

[Stichprobenverfahren] Bei der gezogenen (Brutto-)Stichprobe handelt es sich um eine zufällig gezogene, geschichtete und einstufige Klumpenstichprobe.

Die „primary sampling units“ (Klumpen) wurden anhand der Hochschule und des Studienbereichs⁷ definiert. Die „secondary sampling units“ stellten die Hochschulabsolvent*innen des Prüfungsjahrgangs 2017 innerhalb dieser Klumpen dar. Die Klumpen wurden dabei zunächst nach Abschlussart⁸ und Hochschulart⁹ disproportional geschichtet¹⁰. Innerhalb dieser Schichten wurde dann zufallsbasiert eine Klumpenstichprobe gezogen.

Die Zahl der aus jeder Schicht zu ziehenden Klumpen wurde auf Basis einer Modellierung der durchschnittlich erwarteten Klumpengröße im Prüfungsjahr 2017 ermittelt.¹¹ Die basierte auf der Klumpengröße vorheriger Prüfungsjahre in Relation zur Anzahl an Studienanfänger*innen, Studierenden und Absolvent*innen vorangegangener Studienjahre (Fabian, Flöther & Reifenberg, 2021). Zur Abdeckung der avisierten Stichprobengröße wurden 2.510 Klumpen an 286 Hochschulen gezogen.¹² Die Verantwortlichkeit hinsichtlich der Stichprobenkonzeption und -ziehung lag beim DZHW, während die Akquise der Hochschulen und Administration des Feldzugangs vom ISTAT durchgeführt wurde.

[Rücklauf] Bei Hochschulen, die zeitgleich im Rahmen der vertraglich gebundenen KOAB-Befragung des Prüfungsjahrgangs 2017 eine Vollerhebung aller Absolvent*innen durchführten, erhielten Personen, die einem der gezogenen Stichprobenklumpen angehörten, eine Einladung zur Absolventenbefragung 2017 der Verbundpartner – anstelle der KOAB-Einladung an die restlichen Befragten der Hochschule. Es wird somit zwischen der KOAB-Befragung insgesamt und den Stichprobenklumpen der

⁷ Entsprechend der Aufgliederung nach der amtlichen Statistik (gemäß Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2016/2017 und SoSe 2017).

⁸ Die Ausprägungen waren Bachelor, Master sowie Staatsexamen.

⁹ Die Ausprägungen waren Universität und Fachhochschule.

¹⁰ Aufgrund geringerer Übergangswahrscheinlichkeiten bzw. kleinerer Studienbereiche wurden Masterabsolvent*innen insgesamt sowie Bachelorabsolvent*innen an Universitäten überproportional in die Stichprobe aufgenommen.

¹¹ Zum Zeitpunkt der Stichprobenziehung lagen die amtlichen Zahlen der Hochschulstatistik für das Prüfungsjahr 2017 noch nicht vor.

¹² Hochschulen mit breitem Fächer- und Abschlussspektrum (z.B. „Volluniversitäten“) wiesen in der Regel mehrere Klumpen in der Stichprobe auf, spezialisierte, kleinere Hochschuleschulen – insbesondere theologische, Kunst- und Musikhochschulen – hingegen nur einzelne Klumpen, die stellenweise einer Vollerhebung des Prüfungsjahrgangs entsprachen.

Absolventenbefragung 2017 darin unterschieden. In Hochschulen ohne Bindung an KOAB wurden gezielt nur die Absolvent*innen der dort gezogenen Stichprobenklumpen angesprochen.

Gemäß den Angaben der Verbundpartner unterstützten 111 Hochschulen die Absolventenbefragung 2017. Insgesamt konnten über die Hochschulen 38.365 Personen kontaktiert werden. 12.143 Absolvent*innen nahmen an der Befragung teil (Fabian, Flöther & Reifenberg, 2021). Der editierte Datensatz, welcher dem FDZ-DZHW übergeben wurde, enthielt Daten von 11.447 Befragten. Von diesen mussten weitere 388 Personen ausgeschlossen werden, da diese nicht zur Grundgesamtheit gehörten (stichprobenneutrale Ausfälle: 344 Personen als Promovierte im Prüfungsjahr, 44 Absolvent*innen von Hochschulen außerhalb der Grundgesamtheitsdefinition). Auf dieser Basis ergibt sich näherungsweise eine Netto-Rücklaufquote von 28,83 Prozent.

Weiterführende Informationen zur Kontaktierung der Befragten wie bspw. der Anteil ungültiger Kontaktinformationen, Teilnahmeverweigerungen oder der jeweilige Versandzeitpunkt von Einladungs- und Erinnerungsschreiben liegen dem FDZ-DZHW nicht vor.

4 Datenaufbereitung

Im Folgenden werden die verschiedenen Schritte der Datenaufbereitung beschrieben. Die Generierung von Variablen wurde sowohl durch das Primärforschungsprojekt als auch im Rahmen der Datenedition durch das FDZ-DZHW vorgenommen. Grundlage der im FDZ-DZHW vorgenommenen Datenaufbereitung zur Erstellung des Scientific (SUF) und Campus Use Files (CUF) stellte ein vom ISTAT voreditierter Datensatz dar. Die in den Kapiteln 4.2 und 4.7 dargestellten Tätigkeiten wurden durch das FDZ durchgeführt, teilweise aufbauend auf Vorarbeiten des Primärforschungsprojektes. Die im Rahmen der Datenedition vorgenommenen Aufbereitungsprozesse hinsichtlich der Anonymisierung werden im folgenden Kapitel 5 gesondert erläutert.

Es ist darauf hinzuweisen, dass einige Variablen nicht mehr in ihrer ursprünglichen Form vorhanden waren und die Variablengese teilweise unklar bleibt. Außerdem sind Angaben zu Zeiträumen und Tätigkeitsverläufen in vielen Fällen nicht nutzbar. Hier kommt es bspw. zu fehlenden (Teil-)Angaben, zu unplausiblen Zeitangaben oder Episodenenden, welche vor dem Episodenbeginn liegen. Die entsprechenden Variablen werden im überlieferten Format und unplausibilisiert im SUF oder CUF bereitgestellt. Die meisten zeitbezogenen KOAB-Daten (*Kooperationsprojekt Absolventenstudien* des ISTAT) liegen außerdem als String-Variablen oder im float-Format (nicht in Datumslogik) vor. Dabei handelt es sich insbesondere um die Studien- und Berufsverläufe. Die begrenzte Datenqualität sollte bei etwaigen Analysen berücksichtigt werden. Darüber hinaus gilt es auch zu beachten, dass im Vergleich zu den weiteren DZHW-Absolventenstudienreihen keine Episodendaten zur Verfügung gestellt werden können.

4.1 Datenübertragung

Die Angaben aus der Online-Befragung wurden den DZHW-Verantwortlichen im Primärforschungsprojekt vom Verbundpartner ISTAT als voreditierter Datensatz in Form einer .sav-Datei (SPSS-Datenformat) übermittelt. Einzelne Variablen, die nachgeliefert werden mussten, wurden stellenweise als .xls-Datei zur Verfügung gestellt. Die DZHW-Verantwortlichen übergaben dem FDZ-DZHW einen zusammengeführten Datensatz als .dta-Datei (Stata-Format).

Informationen über Datenformate beim Export aus der Befragungssoftware und der Weiterverarbeitung zum voreditierten Datensatz liegen dem FDZ-DZHW nicht vor.

4.2 Codierung offener Angaben

Die (halb-)offenen Angaben wurden nur in codierter Form in das SUF/CUF aufgenommen. Je nach Variable wurden unterschiedliche Codierlisten verwendet. Es handelt sich um Klassifikationsschlüssel der amtlichen Statistik (z. B. Klassifikation der Berufe, Schlüsselverzeichnis der Studenten- und Prüfungsstatistik etc.) oder um bereits in vorherigen Absolventenpanels eingesetzte Schlüssel. Für einige Variablen wurden neue Codierlisten auf Basis der in den Daten der Absolventenbefragung 2017 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. Einzelne offene Fragen wurden aufgrund von unzureichenden zeitlichen Ressourcen oder weil sie nur als Kontextinformationen zur Codierung anderer offener Angaben erfasst wurden, nicht vercodet. Eine größere Zahl offener Fragen (darunter teilweise auch numerische Abfragen) konnte nicht in das SUF/CUF aufgenommen werden, da vom

Primärforschungsprojekt keine codierte Form dieser Angaben zur Verfügung gestellt wurde und eine zusätzliche Vercodungsleistung über das FDZ-DZHW nicht vorgesehen war.

In 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 – sofern datenschutzrechtlich unbedenklich – im Datensatzreport sowie im Metadaten suchsystem¹³ dokumentiert.

Tabelle 2: Vercodete Merkmale und Codierlisten in der Absolventenbefragung 2017

Merkmal	Codierliste	Codierlisten-ID ^a
Studienfach	Destatis-Fächersystematik 2017/18 [gemäß Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2017/2018 und SoSe 2018), Schlüssel 4.1]	cl-destatis-studienfach-2018
Hochschule	Destatis-Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik, Promovierendenstatistik und Gasthörerstatistik (WiSe 2019/2020 und SoSe 2020), Schlüssel 2.2	cl-destatis-hochschule-2020
Ausland/Staatsangehörigkeit	Projekteigene Referenzliste für Länder (gemäß Destatis), aggr. Länder (gemäß NEPS) und Weltregionen	cl-dzhw-44
Berufsbezeichnung	Destatis-Klassifikation der Berufe 2010	cl-destatis-kladb-2010
sonstige offene Abfragen	Zuordnung zu vorgegebenen Kategorien, projekteigene Codierung oder Löschung	---

a Eine Codierlisten-ID wurde nur dann vergeben, wenn die Kategorien nicht aus den tatsächlichen Nennungen im Datensatz hergeleitet wurden, sondern sich aus einer Systematik ergeben.

4.3 Datenbereinigung

[Löschung von Fällen] In der Absolventenbefragung 2017 wurden - wie bereits oben beschrieben - Fälle aus dem übermittelten Datensatz entfernt. Zunächst wurden 344 Promovierte aus dem Datensatz ausgeschlossen, da diese nicht der definierten Grundgesamtheit angehören¹⁴. Darüber hinaus wurden weitere Fälle (Absolvent*innen von Verwaltungsfachhochschulen und Berufsakademien) identifiziert, die nicht zur Grundgesamtheit gehörten. Diese wurden ebenfalls aus dem Datensatz entfernt.

4.4 Referenzstudium und Studienepisoden

Das Referenzstudium der Befragten bildet eine der zentralen Informationsquellen zur Unterscheidung von Hochschulabsolvent*innen im deutschen Bildungssystem und die Grundlage für eine erheblichen Teil der analytischen Vergleiche. Im Gegensatz zu den DZHW-Absolventenstudienreihen lagen bei der Absolventenbefragung des Jahrgangs 2017 zu den verschiedenen Merkmalen des Referenzstudiums teilweise redundante oder widersprüchliche Informationen vor.

¹³ <https://metadata.fdz.dzhw.eu>

¹⁴ Als Datengrundlage für Promovierte steht das DZHW Promoviertenpanel 2014 zur Verfügung: <https://metadata.fdz.dzhw.eu/en/data-packages/stu-phd2014?page=1&size=10&type=surveys&version=4.0.0>

- Die Angabe zur Hochschule des Referenzstudiums stammt nicht von den Befragten, sondern vom befragenden Verbundpartner auf Basis eines Preloads angespielt.
- Der befragende Verbundpartner arbeitete mit Preloads, auf deren Basis u.a. die Zuordnung der eingeladenen Absolvent*innen in die jeweilige Fragebogenvariante erfolgte.
- Die Preloads dienten ebenfalls zur Einblendung der zugeordneten Abschlussart und Fachrichtung (auf Basis einer partnereigenen Liste) der Befragten zu Beginn des Fragebogens.
- Die Befragten konnte über eine Kontrollschleife sowohl Abschlussart als auch Fachrichtung korrigieren.
- Anfangs- und Enddatum des Studiums sowie die Note wurden in eigenen Items im Fragebogen erhoben.
- Im Rahmen der Befragung wurden auch weitere Studienphasen (nachgelagertes Masterstudium, vorgelagertes Bachelorstudium, Promotionsstudium und weitere Studienphasen) erfasst und als Studienepisoden im Datensatz kumuliert.
- Die teilweise redundanten Informationen (z.B. Preload einer Fachrichtung, Korrektur mit Angabemöglichkeit von bis zu vier Fächern) wurden vom Primärforschungsprojekt stellenweise zusammengeführt und/oder plausibilisiert. Die dort erfolgten Editions Schritte und -entscheidungen waren nur begrenzt rekonstruierbar. Die einzelne Erfassung von Studienphasen (Referenzstudium, Bachelor, Master, Promotion) zeigt dabei stellenweise Abweichungen von den korrespondierende kumulierten Studienepisoden.
- Zur Gewährleistung der Vollständigkeit werden daher im bereitgestellten Datensatz sowohl die kumulierten Studienepisoden als auch die Einzelinformationen zu den spezifischen Studienphasen als Variablen bereitgestellt.
- Im Falle plausibilisierter bzw. editierter Variablen wird zum Erhalt der Übersichtlichkeit für die Nutzenden nur jener Editionsstand verwendet, der nach Abwägung des FDZ-DZHW den jüngsten oder fehlerärmsten Arbeitsstand darstellt.¹⁵ Eine finale Auswahl durch das Primärforschungsprojekt stand nicht zur Verfügung.

Die Entscheidung, welche Studieninformationen für das Referenzstudium oder andere Studienphasen herangezogen werden, obliegt den Nutzenden und dem jeweiligen Analyseinteresse. Es wird darauf hingewiesen, dass sich Studienphasen im Fragebogen und kumulierte Studienepisoden in den bereitgestellten Merkmalen und den jeweiligen Ausprägungen unterscheiden (können).

4.5 Generierung von Variablen

Neben den Variablen, die die codierten Antworten der Befragten enthalten, beinhaltet der Datensatz der Absolventenbefragung 2017 auch generierte Variablen. Dabei handelt es sich zum einen um Variablen mit numerischen Codierungen von ursprünglich offenen Nennungen (vgl. Kapitel 4.2). Zum anderen wurden Variablen hinsichtlich des Datenschutzes verändert (vgl. Kapitel 5) und im Forschungsfeld häufiger benötigte Variablen aus den Werten einer oder mehrerer Quellvariablen generiert. Der Variablenname einer generierten Variablen ist im Datensatz durch das Suffix „_g#“ gekennzeichnet.

Die Dokumentation der einzelnen generierten Variablen mit Angabe ihrer jeweiligen Ausprägungen und Generierungsvorschriften findet sich im Datensatzreport sowie im Metadatensuchsystem¹⁶. Generierte Variablen wurden im Datensatz – sofern möglich – nach der jeweiligen Ausgangsvariable positioniert. Wurde eine Variable aus verschiedenen Quellvariablen generiert, wurde sie hinter jene

¹⁵ Siehe Anhang 1. Dort ist exemplarisch die Variablen mit mögl. redundanten Informationen zum Referenzstudium dargestellt. Grundsätzlich sind sowohl die Angaben der Studienphasen als auch der kumulierten Studienepisoden im Datensatz enthalten. Hierbei kann es jedoch zu Abweichungen voneinander kommen. Dies bedeutet, dass beispielsweise die Angaben zum Promotionsstudium einer*s Befragten im Fragebogenabschnitt Promotion (Studienphase) nicht immer in identischer Form als Studienepisode mit der Abschlussart „Promotion“ vorliegen.

¹⁶ <https://metadata.fdz.dzhw.eu>

Variable eingefügt, die ihr thematisch am nächsten ist. Falls eine eindeutige Zuordnung nicht möglich war, wurde die generierte Variable am Ende des Datensatzes eingefügt.

Außerdem ist zu beachten, dass im Rahmen der Datenaufbereitung im ISTAT weitere Variablen generiert wurden, deren Genese teilweise unklar bleibt.

4.6 Erstellung des Datensatz

[Erstellung des Personendatensatz] Der Personendatensatz enthält den Großteil der nutzbaren Befragungsdaten sowie die zusätzlich generierten Variablen. Pro befragter Person existiert eine Datenzeile (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen. Ein separater Episodendatensatz wird in der Absolventenbefragung 2017 nicht angeboten.

[Dateiformat] Alle Datensätze werden sowohl im Stata- als auch im SPSS-Format bereitgestellt.

4.7 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. Die Variablen- und Wertelabels liegen auf Deutsch vor.

[Variablenbenennung im Personendatensatz] Für die Variablenbenennung wurde im DZHW eine einheitliche Benennungssystematik erstellt, welche sich an den vorherigen Absolventenstudien orientiert. Da jedoch im Rahmen der Absolventenbefragung 2017 eine Variablenbenennung nach dem KOAB-Prinzip gebräuchlich wurde, wurde eine teilweise Anpassung an die DZHW-Systematik durchgeführt. Dabei gilt es zu berücksichtigen, dass die Variablen und deren Benennung nicht deckungsgleich mit den übrigen Kohorten sind. Mit Ausnahme der Identifikatorvariable pid und weiterer Hilfsvariablen des Primärforschungsprojekts wurden die Variablennamen im Personendatensatz nach einem Präfix-Stamm-Suffix-Schema gebildet. Zudem liefern die Variablennamen Metainformationen zur entsprechenden Variable. Das anhand eines Unterstriches vom Stamm abgetrennte Suffix enthält verschiedene Zusatzinformationen, wie die Kenntlichmachung von generierten Variablen.

Bestimmte Variablen sind aus Anonymisierungsgründen nicht über alle potentiellen Zugangswege (Download-CUF, Download-SUF, Remote-Desktop-SUF, On-Site-SUF; s. Kapitel 5) einsehbar. In diesen Fällen wird bei den Personendaten im Variablenlabel in Klammern derjenige Zugangsweg angegeben, ab dem die Variable nutzbar ist:

- c: Variable ist im CUF, im Download-SUF, im Remote-Desktop-SUF und im On-Site-SUF nutzbar.
- d: Variable ist nicht im CUF, aber im Download-SUF, im Remote-Desktop-SUF und im On-Site-SUF nutzbar.
- r: Variable ist nicht im CUF und im Download-SUF, aber im Remote-Desktop-SUF und im On-Site-SUF nutzbar.
- o: Variable ist nicht im CUF, im Download-SUF und im Remote-Desktop-SUF, aber im On-Site-SUF nutzbar.
- a: Variable ist über keinen Zugangsweg nutzbar. Sie wird aber dokumentiert, da es zugehörige Fragen im Fragebogen gibt.

4.8 Codierung fehlender Werte

Die Codierung fehlender Werte wurde vom Primärforschungsprojekt übernommen und aufgrund der vorliegenden Filterführung nicht verändert oder angepasst. Die Systematik gestaltet sich wie folgt:

Tabelle 3: Systematik des Primärforschungsprojekts für fehlende Werte

Code	Wertelabel
-9	keine Angabe
-8	Item non-response
-7	gefiltert
-6	nicht bestimmbar
-5	gefiltert und dropout
-4	dropout
-3	unplausibler Wert
-2	mfn im Papierfragebogen ¹⁷
-1	Frage nicht gestellt
-967	anonymisiert

¹⁷ Hierbei handelt es sich um Mehrfachnennungen bei Fragen mit Einfachauswahl in Papierfragebögen (wird in Online-Befragungen technisch verhindert). Da die Absolventenbefragung 2017 als Online-Befragung erfolgte, spielt dieser fehlende Wert hier keine Rolle.

5 Anonymisierung

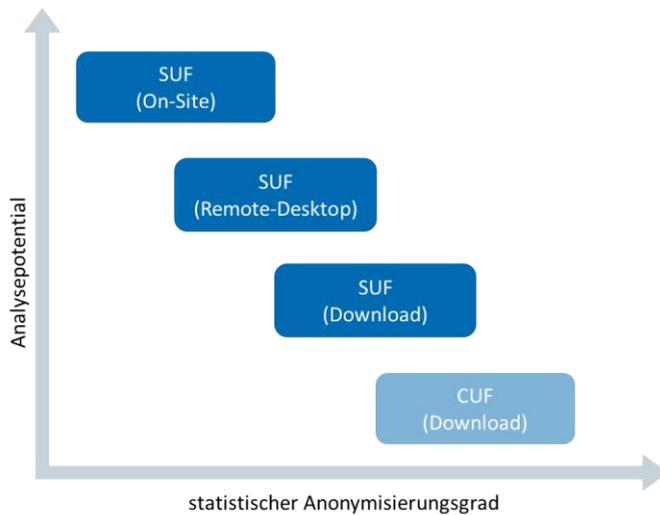
[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten¹⁸, die in freiwilligen Befragungen durch das DZHW erhoben werden, gelten die EU-Datenschutz-Grundverordnung (EU-DSGVO) und das Bundesdatenschutzgesetz. Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (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] Das FDZ-DZHW stellt für die Absolventenbefragung 2017 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 jeden Zugangsweg wird eine andere SUF-Variante bereitgestellt, die unterschiedlich stark anonymisiert worden ist und entsprechend mehr oder weniger Informationen umfasst. Abbildung 2 gibt einen Überblick über den jeweiligen Grad der statistischen Anonymisierung und das damit verbundene Analysepotential. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen in Abhängigkeit von Datenprodukt (SUF/CUF) und Zugangsweg erläutert.

¹⁸ „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).

Abbildung 2: Datenzugangswege, statistischer Anonymisierungsgrad und Analysepotential der Daten der Absolventenbefragung 2017



[Statistische Anonymisierungsmaßnahmen] Im Rahmen der Anonymisierung wurden zunächst alle Informationen, mit denen sich Personen oder Institutionen direkt identifizieren lassen, gelöscht. Diese sogenannten *direkten Identifikatoren* wie Namen, Adressen und E-Mail-Adressen wurden in der Absolventenbefragung 2017 bereits während der Feldphase in einem separaten Datensatz erfasst (vgl. Kapitel Fehler! Verweisquelle konnte nicht gefunden werden.) und sind somit weder im CUF noch in den verschiedenen SUF-Varianten enthalten. Um einen Rückbezug auf diesen Datensatz zu unterbinden, wurde zudem die Original-Identifikationsnummer 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.¹⁹ Für die Absolventenbefragung 2017 wurden die folgenden Quasi-Identifikatoren identifiziert, die sowohl in externen Datenquellen²⁰ als auch in den Absolvent*innendaten vorliegen: Hochschule, Studienfach, Abschlussart, Promotions- und Berufsangaben, regionale Informationen (zur Hochschule, zum Ort des Erwerbs der Studienberechtigung, Arbeitsort und eines Auslandsaufenthalts), Staatsangehörigkeit und Geburtsland. Um eine eindeutige Zuordnung der Absolvent*innendaten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Datenprodukt bzw. Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 4Tabelle 4). Beispielsweise wird das Merkmal „Hochschule“ in dem SUF für die On-Site Nutzung zu NUTS-2-Regionen, im Remote-Desktop-SUF zu Bundesländern, im Download-SUF und im Download-CUF zu den zwei Kategorien alte vs. neue Bundesländern aggregiert. Offene Angaben sind ebenfalls Quasi-Identifikatoren (vgl. Ebel, 2015, S. 3) und wurden im Rahmen der Anonymisierung vercodet oder gelöscht.

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 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 Absolventenbefragung 2017 wurden Gesundheitsinformationen erhoben, für die bei den Befragten kein zusätzliches Einverständnis für die

¹⁹ 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.

²⁰ z.B. Studenten- und Prüfungsstatistik des statistischen Bundesamtes, Alumninetzwerke der Hochschulen oder auch Berufsnetzwerke.

Sekundärnutzung eingeholt wurde. Daher wurden diese Antworten im CUF- und allen SUF-Varianten gelöscht.

Zur Gewährleistung der absoluten Anonymisierung der Daten des CUF wurden zum einen im Vergleich zu den SUF-Varianten restriktivere statistische Anonymisierungsmaßnahmen auf Variablenebene vorgenommen (vgl. Tabelle 4). Zum anderen wurde eine per Zufallsauswahl gewonnene 2/3-Substichprobe der Daten gezogen.

Tabelle 4: Maßnahmen der statistischen Anonymisierung der Daten der Absolventenbefragung 2017 nach Zugangsweg

Merkmal	On-Site-SUF	Remote-Desktop-SUF	Download-SUF	Download-CUF (Substichprobe)
Direkte Identifikatoren	Löschung und Vergabe einer zufälligen ID			
Fragebogeneingang	Freigabe	Freigabe von Monat/Jahr	Freigabe von Monat/Jahr	Freigabe von Jahr
Studienfach	Vercodung zu Studienfächern ^a	Aggregation zu Studienbereichen ^a	Aggregation zu Studienbereichen ^a	Aggregation zu Fächergruppen ^a
Abschlussart	Aggregation zu Kategorien: - Bachelor (ohne Lehramt) - Master (ohne Lehramt) - Staatsexamen (ohne Lehramt) - Bachelor (Lehramt) - Master (Lehramt) - Staatsexamen (Lehramt) - Lehramt ohne nähere Angabe - Promotion - sonstiger Abschluss	Aggregation zu Kategorien: - Bachelor (ohne Lehramt) - Master (ohne Lehramt) - Staatsexamen (ohne Lehramt) - Bachelor (Lehramt) - Master (Lehramt) - Staatsexamen (Lehramt) - Lehramt ohne nähere Angabe - Promotion - sonstiger Abschluss	Aggregation zu Kategorien: - Bachelor (ohne Lehramt) - Master (ohne Lehramt) - Staatsexamen (ohne Lehramt) - Bachelor (Lehramt) - Master (Lehramt) - Staatsexamen (Lehramt) - Lehramt ohne nähere Angabe - Promotion - sonstiger Abschluss	Aggregation zu Kategorien: - Bachelor (ohne Lehramt) - Master (ohne Lehramt) - Staatsexamen (ohne Lehramt) - Bachelor (Lehramt) - Master (Lehramt) - Staatsexamen (Lehramt) - Lehramt ohne nähere Angabe - Promotion - sonstiger Abschluss
Hochschule	Aggregation zu Hochschulart (6 Kategorien) und Hochschulort- (NUTS-2-Regionen) ^b	Aggregation zu Hochschulart (6 Kategorien) und Hochschulort- (Bundesländer)	Aggregation zu Hochschulart (Uni/FH) und Hochschulort (neue bzw. alte Bundesländer)	Aggregation zu Hochschulart (Uni/FH) und Hochschulort (neue bzw. alte Bundesländer)
Datumsangaben Referenzstudium	Freigabe Quartal (Aggregation von Monaten)/Jahr	Freigabe Quartal (Aggregation von Monaten)/Jahr	Freigabe Quartal (Aggregation von Monaten)/Jahr	Löschung
Erwerb der Studienberechtigung: Erweiterte Oberschule (DDR) und Fach- und Ingenieurschule	Löschung	Löschung	Löschung	Löschung
Zeitpunkt Erwerb Studienberechtigung	Löschung	Löschung	Löschung	Löschung
Ort des Erwerbs der Studienberechtigung (Bundesland/ Ausland)	Freigabe	Freigabe	Freigabe	Aggregation zu alte/neue Bundesländer und Ausland

Ort des Erwerbs der Studienberechtigung (PLZ)	Freigabe	Aggregation zu NUTS-2-Regionen ^b	Aggregation zu NUTS-2-Regionen ^b	Löschung
Zeitpunkt Abschluss der Ausbildung	Freigabe	Aggregation zu den Kategorien: - „1968-1997“, - „2015-2018“, der Rest wird einzeln ausgewiesen.	Aggregation zu den Kategorien: - „1968-1997“, - „2015-2018“, der Rest wird einzeln ausgewiesen.	Aggregation zu den Kategorien: - „1968-1997“, - „2015-2018“, der Rest wird einzeln ausgewiesen.
Abschluss der Aufstiegsfortbildung	Freigabe	Aggregation zu den Kategorien: - „1987-2007“, - „2015-2018“, der Rest wird einzeln ausgewiesen	Aggregation zu den Kategorien: - „1987-2007“, - „2015-2018“, der Rest wird einzeln ausgewiesen	Aggregation zu den Kategorien: - „1987-2007“, - „2015-2018“, der Rest wird einzeln ausgewiesen
Datenübermittlung an die Studienstiftung des deutschen Volkes	Löschung	Löschung	Löschung	Löschung
Wohnsitz und Arbeitsort (PLZ)	Freigabe	Aggregation zu NUTS-2-Regionen ^b	Aggregation zu NUTS-2-Regionen ^b	Löschung
Wohnsitz und Arbeitsort (Ausland)	Aggregation nach NEPS Klassifikation ^d	Aggregation nach NEPS Klassifikation ^d	Aggregation zu Weltregionen	Löschung
Beruf	Vercodung zu KldB-5-Steller ^c	Aggregation zu KldB-3-Steller ^c	Aggregation zu KldB-3-Steller ^c	Aggregation zu KldB-2-Steller ^c
Unternehmensgröße	Aggregation zu den Kategorien: - „1-49 Beschäftigte“ - „50-249 Beschäftigte“ - „250-999 Beschäftigte“ - „über 1000 Beschäftigte“	Aggregation zu den Kategorien: - „1-49 Beschäftigte“ - „50-249 Beschäftigte“ - „250-999 Beschäftigte“ - „über 1000 Beschäftigte“	Aggregation zu den Kategorien: - „1-49 Beschäftigte“ - „50-249 Beschäftigte“ - „250-999 Beschäftigte“ - „über 1000 Beschäftigte“	Aggregation zu den Kategorien: - „1-49 Beschäftigte“ - „50-249 Beschäftigte“ - „250-999 Beschäftigte“ - „über 1000 Beschäftigte“
Angaben zu Sonderzahlungen, Zulagen und einem Einkommen > 6.500€	Löschung	Löschung	Löschung	Löschung
Geburtsjahr	Freigabe	Freigabe	Aggregation zu den Kategorien: - „1976 und jünger“, - „1996+1997“, der Rest wird einzeln ausgewiesen	Aggregation zu den Kategorien: - „1976 und jünger“, - „1996+1997“, der Rest wird einzeln ausgewiesen
Geschlecht	Kategorie „divers“ wird nicht freigegeben			
Geburtsmonat und -jahr Kinder	Freigabe	Geburtsmonat: Löschung Geburtsjahr: Aggregation	Geburtsmonat: Löschung Geburtsjahr: Aggregation	Geburtsmonat: Löschung Geburtsjahr: Aggregation
Angaben zur Staatsangehörigkeit	Freigabe	Aggregation nach NEPS Klassifikation ^d	Aggregation zu Weltregionen	Löschung
Angaben zu Geburtsländern	Freigabe	Aggregation nach NEPS Klassifikation ^d	Aggregation zu Weltregionen	Löschung
Jahr der Einwanderung	Freigabe	Freigabe	Freigabe	Löschung
Merkmale zur Gesundheit	Löschung	Löschung	Löschung	Löschung
sonstige offene Angaben	Löschung	Löschung	Löschung	Löschung

- a Nach dem Schlüsselverzeichnis der Studenten- und Prüfungsstatistik WiSe 2019/2020 und SoSe 2020 von Destatis.
- b Statistisches Amt der Europäischen Union (Eurostat): Nomenclature of territorial units for statistics (NUTS) <http://ec.europa.eu/eurostat/web/nuts/overview>.
- c Nach der Klassifikation der Berufe von 2010 von Destatis.
- d Die Aggregation der Staaten zu Weltregionen ist angelehnt an die Klassifikation des NEPS https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/6-0-0/SC5_6-0-0_Anonymisation.pdf (in der ersten Welle mit Anpassungen bei europäischen Ländern).

6 Literatur

- Ebel, T. (2015). Empfehlungen zur Anonymisierung quantitativer Daten. Mannheim: GESIS- Leibniz-Institut für Sozialwissenschaften.
- Fabian, G., Flöther, C. & Reifenberg, D. (2021). Generation Hochschulabschluss: neue und alte Differenzierungen. Ergebnisse des Absolventenpanels 2017. Münster, New York: Waxmann.
- Hoffstätter, U., Niebuhr, J., & Vietgen, S. (2021). *DZHW-Absolventenpanel 2013. Daten- und Methodenbericht zur Absolvent*innenkohorte 2013 (1. und 2. Befragungswelle)*. Hannover: FDZ-DZHW.
- 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
- Rammstedt, B. (2007). The 10-Item Big Five Inventory. Norm values and investigation of sociodemographic effects based on a German population representative sample. *European Journal of Psychological Assessment*, 23(3), 193–201. doi:10.1027/1015-5759.23.3.193
- Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less. A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212. doi:10.1016/j.jrp.2006.02.001 34 D Z H W-A b s o l v e n t e n p a n e l 2 0 0 9
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2013). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit. *10 Item Big Five Inventory (BFI-10). methoden, daten, analysen*, 7(2), 233–249. doi:10.12758/mda.2013.013
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C. & Kovaleva, A. (2014). Big Five Inventory (BFI10). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis76

Anhang 1

Varname	Beschreibung	Empfehlung
Kontrollschleife		
b1pjvalid	<p>Abfrage der „korrekten“ Zuordnung.</p> <p>Ermöglicht den Befragten, die Zuordnung einer Abschlussart und einer Fachrichtung aus den Preloads zu bestätigen oder Abweichungen dieser beiden Merkmale zu benennen.</p> <p>Im Primärforschungsprojekt wurden verschiedene „korrekte“ Variablen aus der Kontrollschleife generiert, die je nach Verbundpartner voneinander abwichen.</p> <p>Es ist nicht ersichtlich, ob spätere Einblendungen der Preloads im Fragebogen die hier gemachten Korrekturen berücksichtigen.</p>	
Abschlussart		
b1pjabs_g1	Durch das Primärforschungsprojekt „korrigierte“ Abschlussart – bereitgestellt durch das Primärforschungsprojekt. Durch das FDZ-DZHW aus Studiengängen auf Abschlüsse aggregiert.	<p>Soll das Referenzstudium als Studienphase abdecken.</p> <p>Redundanz mit b1hfabsp1_g1 beachten.</p>
b1hfabsp1_g1	<p>Erste kumulierte Studienepisode. Laut Labelung des übermittelten Datensatzes soll es sich dabei um das Referenzstudium handeln.</p> <p>Allerdings zeigten sich geringe Abweichungen zwischen dieser Variable und der korrigierten Abschlussart in b1jabs_g1, insbesondere durch Missings in einer Variable, die in der anderen gültige Werte aufweist. Auch innerhalb der aggregierten gültigen Angaben bestehen noch bei ca. 0,5% der Fälle Abweichungen.</p> <p>Um Irritationen zu vermeiden, wurde die Variable als „Abschlussart Studium Phase 1“ gelabelt, um sie von der Alternative b1pjabs_g1 für die Abschlussart des Referenzstudium zu unterscheiden.</p>	Ob die kumulierten Studienepisoden oder die Angaben der Studienphasen für das Referenzstudium verwendet werden, obliegt der Entscheidung der Nutzenden.
b1hfabsp2_g1 bis b1hfabsp7_g1	<p>Abschlussarten der weiteren Studienepisoden. Unklar ist, ob tlw. die späteren Studienepisoden das „wahre“ Referenzstudium umfassen, wenn Befragte bei der Referenzepisode außerhalb des Prüfungsjahrs liegen.</p> <p>Diese Frage besteht insbesondere, da ein nennenswerter Anteil der Befragten ein Enddatum der ersten Studienepisode außerhalb der Jahre 2016 und 2017 (den Kalenderjahren des Prüfungsjahrs), einige davon dann aber in den späteren Studienepisoden Enddaten im Jahr 2017 angeben.</p>	Ob die kumulierten Studienepisoden oder die Angaben der Studienphasen verwendet werden, obliegt der Entscheidung der Nutzenden. Gilt sowohl für das Referenzstudium, das nachgelagerte Masterstudium, das vorgelagerte Bachelorstudium, als auch eine Promotion.
Studienfach		
b1pjhfach_p1	Studienfach nach Destatis Fächersystematik WS 2017/18.	Trotz gewisser Unsicherheit aufgrund mehrerer potentieller Variablen

	Generiert vom Primärforschungsprojekt. Im Rahmen der Erhebung wurde die Fachrichtung über eine partnereigene Liste abgefragt. Im Primärforschungsprojekt wurden verschiedene „korrekte“ Variablen aus der Kontrollschleife generiert, die je nach Verbundpartner voneinander abwichen.	zur Fachrichtung im übermittelten Datensatz, nach Einschätzung des FDZ-DZHW die geeignetste Variable.
b1pjhfach_g1	Studienbereich nach Destatis Fächersystematik WS 2017/18. Generiert vom Primärforschungsprojekt. Im Rahmen der Erhebung wurde die Fachrichtung über eine partnereigene Liste abgefragt. Im Primärforschungsprojekt wurden verschiedene „korrekte“ Variablen aus der Kontrollschleife generiert, die je nach Verbundpartner voneinander abwichen.	Trotz gewisser Unsicherheit aufgrund mehrerer potentieller Variablen zur Fachrichtung im übermittelten Datensatz, nach Einschätzung des FDZ-DZHW die geeignetste Variable für das Referenzstudium.
b1pjhfach_g2	Fächergruppe nach Destatis Fächersystematik WS 2017/18. Generiert vom Primärforschungsprojekt. Im Rahmen der Erhebung wurde die Fachrichtung über eine partnereigene Liste abgefragt. Im Primärforschungsprojekt wurden verschiedene „korrekte“ Variablen aus der Kontrollschleife generiert, die je nach Verbundpartner voneinander abwichen.	Trotz gewisser Unsicherheit aufgrund mehrerer potentieller Variablen zur Fachrichtung im übermittelten Datensatz, nach Einschätzung des FDZ-DZHW die geeignetste Variable für das Referenzstudium.
b1hfachp1	Erste kumulierte Studienepisode. Laut Labelung des übermittelten Datensatzes soll es sich dabei um das Referenzstudium handeln. Aufgrund fehlender Verkodung lässt sich eine Redundanz mit b1pjhfach_p1 nicht überprüfen.	Aufgrund fehlender Verkodung keine Verwendung möglich.
b1hfachp2 bis b1hfachp7	Erste kumulierte Studienepisode. Ebenfalls keine Verkodung vorhanden. Diese Studienepisoden können identisch mit den Studienphasen (Referenzstudium, Master, Bachelor, Promotion) sein. Dies kann aber nicht abschließend beantwortet werden.	Aufgrund fehlender Verkodung keine Verwendung möglich.
Zeiten (Anfang/Ende)		
b1pjjanf_g1 und b1pjqanf_g1	Anfangjahr und -quartal (aggregiert aus Monat) des Referenzstudiums. Genese nicht eindeutig, da Zeitangaben der Studienphase Referenzstudium zu großen Teilen identisch mit Zeitangaben der kumulierten ersten Studienepisode, aber dennoch mit Abweichungen behaftet.	Weist gegenüber den Studienepisoden mehr gültige Angaben auf. Nach Einschätzung des FDZ-DZHW die geeignetste Variable für das Referenzstudium.
b1pjwend_g1 und b1pjwend_g1	Endjahr und -quartal (aggregiert aus Monat) des Referenzstudiums. Genese nicht eindeutig, da Zeitangaben der Studienphase Referenzstudium zu großen Teilen identisch mit Zeitangaben der kumulierten ersten Studienepisode, aber dennoch mit Abweichungen behaftet.	Weist gegenüber den Studienepisoden mehr gültige Angaben auf. Nach Einschätzung des FDZ-DZHW die geeignetste Variable für das Referenzstudium.
b1janfp1 und b1janfp1_g1	Anfangsjahr erste kumulierte Studienepisode. Laut Label Beginn Referenzstudium, allerdings weichen b1janfp1 und b1pjjanf_g1 bei ca. 300 Befragten voneinander ab.	Ob die kumulierten Studienepisoden oder die Angaben der Studienphasen verwendet werden, obliegt der Entscheidung der Nutzenden.
b1jendp1 und b1jendp1_g1	Endjahr erste kumulierte Studienepisode. Laut Label Beginn Referenzstudium, allerdings weichen b1jendp1 und b1pjwendf_g1 bei einigen Befragten voneinander ab.	Ob die kumulierten Studienepisoden oder die Angaben der Studienphasen verwendet werden, obliegt der

		Entscheidung der Nutzenden.
b1janfp2 und b1janfp2_g1 bis b1janfp7 und b1janfp7_g1; b1jendp2 und b1jendp2_g1 bis b1jendp7 und b1jendp7_g1	<p>Anfangs- und Enddaten der weiteren kumulierten Studienepisoden. Es kann sich dabei um Redundanzen zu den jeweiligen Studienphasen (Referenzstudium, Master, Bachelor, Promotion) handeln. Dies konnte allerdings auf Basis der vorliegenden Datumangaben nicht überprüft werden.</p> <p>Wie auch bei anderen Merkmalen weichen die Variablen der Studienphasen und -episoden bei den vom Primärforschungsprojekt erzeugt Missing-Kategorien teilweise deutlich voneinander ab.</p>	Ob die kumulierten Studienepisoden oder die Angaben der Studienphasen verwendet werden, obliegt der Entscheidung der Nutzenden.
Abgeschlossen Prüfung/Status		
b1pjstand	<p>Statusvariable für das Referenzstudium. Genese durch das Primärforschungsprojekt unklar. Scheint eine aggregierte Form von b1standp1_g1 zu sein.</p> <p>Allerdings weichen die Angaben zwischen beiden Variablen tlw. voneinander ab, insbesondere bei den Missingkategorien.</p>	Ob die kumulierten Studienepisoden oder die Angaben der Studienphasen verwendet werden, obliegt der Entscheidung der Nutzenden.
b1standp1_g1	Statusangabe der ersten kumulierten Studienepisode. Lauf Label im übermittelten Datensatz soll es sich um das Referenzstudium handeln, allerdings bestehen Abweichungen gegenüber b1pjstand.	Ob die kumulierten Studienepisoden oder die Angaben der Studienphasen verwendet werden, obliegt der Entscheidung der Nutzenden.
b1standp2_g1 bis b1standp7_g1	<p>Statusangabe der weiteren kumulierten Studienepisoden. Es kann sich dabei um Redundanzen zu den jeweiligen Studienphasen (Referenzstudium, Master, Bachelor, Promotion) handeln.</p> <p>Wie auch bei anderen Merkmalen weichen die Variablen der Studienphasen und -episoden bei den vom Primärforschungsprojekt erzeugt Missing-Kategorien teilweise deutlich voneinander ab.</p>	Ob die kumulierten Studienepisoden oder die Angaben der Studienphasen verwendet werden, obliegt der Entscheidung der Nutzenden.
Hochschule		
b1pjhs	<p>Anders als bei Abschlussart, Fachrichtung, Status und Zeiten ist die Hochschule nicht in den kumulierten Studienepisoden, sondern nur in den Studienphasen ausgeführt. Die Hochschule des Referenzstudiums wird an keiner Stelle erhoben, so dass davon auszugehen ist, dass diese als Preload vom Primärforschungsprojekt angespielt wurde.</p> <p>Wurde vom Primärforschungsprojekt in Destatis Hochschulcodes verkodet. Variable bildet die Grundlage für alle Aggregationen der Hochschule.</p>	Keine Redundanz, normale Verwendung.