

Henrike Schmidtchen/Robert Birkelbach

DZHW-Absolventenpanel 1993

Daten- und Methodenbericht zu den Erhebungen der
Absolvent*innenkohorte 1993 (1. und 2. Befragungswelle)

Daten- und Methodenbericht

Dezember 2020

Autor/in:
Henrike Schmidtchen
Robert Birkelbach

Ein besonderer Dank gilt Anne Weber und Andreas Daniel sowie dem DZHW Kollegen Christian Kerst für die hilfreiche Unterstützung.

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

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

Vorsitzender des Aufsichtsrats:
Ministerialdirigent Peter Greisler

Registergericht:
Amtsgericht Hannover | B 210251

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



Inhaltsverzeichnis

Tabellenverzeichnis	IV
Datennutzungshinweise	5
1 Inhalt und Anlage der Studie	7
2 Grundgesamtheit und Stichprobenverfahren	9
3 Design, Durchführung und Rücklauf	10
4 Datenaufbereitung	11
4.1 Personendaten	11
4.2 Episodendaten	12
5 Anonymisierung	15
6 Literaturverzeichnis	19
Anhang	20
A. Referenzliste: Fragen, die in beiden Befragungswellen vorkommen	20

Tabellenverzeichnis

Tabelle 1:	Anlage der Hauptuntersuchungen der DZHW-Absolventenstudienreihe von 1989 und 1993	7
Tabelle 2	Systematik für fehlende Werte.....	12
Tabelle 3:	Zuordnungsschema Tätigkeitsarten	13
Tabelle 4:	Maßnahmen der statistischen Anonymisierung nach Zugangsweg	17

Datennutzungshinweise

[Voraussetzungen der Datennutzung] Die Daten des Absolventenpanels 1993 werden durch das FDZ des DZHW entsprechend der europäischen Datenschutzgrundverordnung (EU-DSGVO) anonymisiert und als Scientific Use File (SUF) ausschließlich zur wissenschaftlichen Sekundärnutzung freigegeben.¹ Das SUF kann über die Website des FDZ (<https://fdz.dzhw.eu>) beantragt werden. Für die Nutzung wird ein Datennutzungsvertrag abgeschlossen.

[Datenzugang] Das SUF wird über zwei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ-DZHW 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.
- **On-Site:** Die Daten werden in den Räumlichkeiten des FDZ in einer kontrollierten Umgebung an einem speziell gesicherten Computer zur Analyse bereitgestellt. Wie beim Remote-Desktop-Zugang werden Analyseergebnisse erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.

Die bereitgestellten Daten weisen je nach Zugangsweg einen unterschiedlich hohen Informationsgehalt auf und unterscheiden sich damit hinsichtlich ihres Analysepotentials. 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.² Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

[Dokumentationsmaterialien] Der vorliegende Daten- und Methodenbericht ist ein Part der Dokumentation zur ersten und zweiten Befragungswelle des Absolventenpanels 1993. Alle weiteren Dokumentationsmaterialien (Datensatzreports, Fragebögen, etc.) zur Studie können frei im Metadatensuchsystem des FDZ-DZHW unter [10.21249/DZHW:gra1993:1.0.0](#) heruntergeladen werden.

¹ Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, 6ff.), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LIbBi) (vgl. Koberg, 2016, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker & Roß, 2012, 9f.) 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: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen.

² Zu den verschiedenen Anonymisierungsgraden und Analysepotentialen der beiden SUF-Varianten vgl. Kapitel 5.

[Gebühren der Datenbereitstellung] SUF werden derzeit (Stand: Dezember 2020) kostenfrei zur Verfügung gestellt. Änderungen bzw. das aktuelle Kostenmodell können auf der Website des FDZ-DZHW (<https://fdz.dzhw.eu>) eingesehen werden.

[Pflichten der Datennutzer*innen] Die Datennutzer*innen sind verpflichtet, folgende Regeln³ 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 des Datenschutzes bzw. der Datensicherheit erlangen, müssen diese dem FDZ-DZHW unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** Die Daten dürfen nur durch die Personen genutzt werden, die einen Datennutzungsvertrag abgeschlossen haben.
- **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.

³ Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

1 Inhalt und Anlage der Studie

[Studienreihe] Die DZHW-Absolventenstudien sind eine Untersuchungsreihe, in der anhand von standardisierten Befragungen Informationen zu Studium, Berufseintritt, Berufsverlauf und Weiterqualifizierung von Hochschulabsolvent*innen erfasst werden. Sie werden durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)⁴ durchgeführt, vom Bundesministerium für Bildung und Forschung (BMBF) gefördert und dienen – in Ergänzung zur amtlichen Hochschulstatistik – dem nationalen Bildungsmonitoring.

Beim Absolventenpanel 1993 handelt es sich um die zweite Absolvent*innenkohorte dieser Studienreihe. Seit dem Absolventenpanel 1989 wird jeder vierte Absolvent*innenjahrgang (Kohorte) untersucht. Die Grundgesamtheit einer Kohorte sind Hochschulabsolvent*innen, die im Winter- oder Sommersemester des betreffenden Prüfungsjahrs ihren Studienabschluss an einer Hochschule in der Bundesrepublik Deutschland erworben haben.

Für jede Absolvent*innenkohorte werden mehrere Befragungswellen durchgeführt. Die Befragungen der Absolvent*innenkohorten von 1989 und 1993 umfassten zwei Wellen, seit 1997 wird eine zusätzliche dritte Befragungswelle durchgeführt. Jede Welle findet in unterschiedlichen zeitlichen Abstand zum Studienabschluss statt. Die erste Welle einer Kohorte findet im Mittel ein bis eineinhalb Jahre nach dem jeweiligen Studienabschluss statt. Die zweite Befragungswelle folgt etwa fünf Jahre und die dritte Befragungswelle etwa zehn Jahre nach dem Studienabschluss. Der thematische Fokus einer Befragungswelle orientiert sich an der jeweils typischen Bildungs-, Berufs- und Lebensphase der Befragten zum Befragungszeitpunkt.

Die verschiedenen Erhebungen in den Absolvent*innenjahrgängen 1989 und 1993 wurden als schriftlich-postalische Paper-and-Pencil-Befragung (Paper and Pencil Interview, PAPI) durchgeführt. Im Rahmen der jüngeren Absolventenpanel wird jedoch zunehmend auf Online-Befragungen (Computer Assisted Web Interview, CAWI) zurückgegriffen.

Tabelle 1: Anlage der Hauptuntersuchungen der DZHW-Absolventenstudienreihe von 1989 und 1993

Welle	Befragungszeitpunkt	Thematischer Schwerpunkt
1	ca. 1 Jahr nach Studienabschluss	Studienverlauf und -erfahrung, akademische Weiterqualifizierung, Übergang in den Beruf
2	ca. 5 Jahre nach Studienabschluss	Aktuelle Tätigkeit, Erwerbstätigkeit, akademische und berufliche Weiterbildung

[Analysepotential] In allen Kohorten wird ein identischer Kern an Informationen erhoben (s. Tabelle 1). Auf dieser Grundlage können anhand von Zeitreihen- bzw. Kohortenanalysen langfristige Trends der Hochschulbildung und Arbeitsmarktentwicklung betrachtet werden. Zudem wird innerhalb einer Kohorte ein Teil der Fragen in den verschiedenen Befragungswel-

⁴ Das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW, <http://www.dzhw.eu>) entstand im August 2013 durch eine Ausgründung aus der HIS Hochschul-Informationssystem GmbH. Im nachfolgenden Text wird durchgängig der Begriff DZHW verwendet, auch wenn die Studie vor der Ausgründung 2013 durchgeführt wurde.

len wiederholt gestellt. Dies ermöglicht die Betrachtung intra-individueller Veränderungen zwischen den Wellen (z. B. kausale Panelanalysen). Besonders hervorzuheben ist, dass in allen Kohorten über die Wellen hinweg monatsgenaue und kontinuierliche Verlaufsdaten zum individuellen Tätigkeitsverlauf seit dem Studienabschluss erfasst werden, die sich für Ereignisdaten- und Sequenzmusteranalysen eignen. Darüber hinaus werden in Abhängigkeit von aktuell-aktuellen Entwicklungen und Forschungsinteressen in einzelnen Kohorten bestimmte Aspekte vertiefend oder ergänzend abgefragt.

[Spezifika des Absolventenpanels 1993] Neben den allgemeinen Charakteristika der Studienreihe ist bzgl. des Absolventenpanels 1993 besonders hervorzuheben, dass die Hochschulabsolvent*innen des untersuchten Prüfungsjahrgangs mit gravierenden Berufsstartänderungen konfrontiert wurden. Diese ergeben sich aus der 1991 eingetretenen Rezession, welche verschlechterte Berufschancen sowie Veränderungen in Einsatzfeldern, die auf Tätigkeiten in der güterproduzierenden Industrie gerichtet sind, zur Folge hatte (Holtkamp, Minks & Schaeper, 1998). Aufgrund dieser teils krisenhaften Berufsstartphasen der Absolvent*innen 1993 vor allem technisch-naturwissenschaftlicher Studiengänge wurden neben Veränderungen in den Strukturen und Bedingungen des Berufseinstiegs insbesondere die berufliche Integration und Desintegration untersucht.

Befragt wurde der Prüfungsjahrgang 1992/1993, wobei erstmals nach der deutschen Wiedervereinigung 1990 Absolvent*innen aus allen Bundesländern der Bundesrepublik Deutschland ausgewählt wurden.

2 Grundgesamtheit und Stichprobenverfahren

[Grundgesamtheit] Die Grundgesamtheit des Absolventenpanels 1992 umfasst alle Hochschulabsolvent*innen, die im Wintersemester 1992/1993 oder im Sommersemester 1992 ihren ersten berufsqualifizierenden Studienabschluss an einer staatlich anerkannten Hochschule in der Bundesrepublik Deutschland erworben haben (mit Ausnahme der Absolvent*innen von Verwaltungsfachhochschulen). Sie umfasst 162.611 Absolvent*innen (Rehn, Brandt, Fabian & Briedis, 2011, S. 3).

[Geschichtete Klumpenstichprobe] Aufgrund fehlender oder nicht zugänglicher Listen von Hochschulabsolvent*innen mussten die Individuen über die Hochschulen rekrutiert werden. Dabei wurde eine geschichtete Klumpenstichprobe gezogen. Die „primary sampling units“ (Klumpen) wurden anhand der Hochschule, des Studienbereichs⁵ und der Abschlussart⁶ definiert. Die Umsetzung dieses Designs erfolgte über die Ziehung von einfachen Zufallsstichproben. Die Klumpen wurden dabei nach Hochschulart (Uni/FH), regionaler Verteilung der Studienangebote und Größenstruktur der Hochschule geschichtet, wobei eine detaillierte Differenzierung der letzten beiden Merkmale nicht mehr rekonstruierbar ist. In den jeweiligen Schichten der Stichproben bestand das Ziel, eine im Vergleich zur Grundgesamtheit proportionale Verteilung zu erreichen. Die „secondary sampling units“ stellten die Hochschulabsolvent*innen des Prüfungsjahrgangs 1993 innerhalb der Klumpen dar (die Klumpen wurden jeweils vollständig erhoben).

Zur Umsetzung dieses Stichprobenverfahrens wurden die gezogenen Hochschulen vom DZHW angeschrieben und um ihre Teilnahme gebeten. Zudem teilte das DZHW den Hochschulen bzw. deren Prüfungsämtern die Kriterien (Prüfungsjahr, Erstabschluss, Studienbereich, Abschlussart) mit, anhand derer sie die Zielpersonen für das Absolventenpanel 1993 identifizieren sollten.

⁵ nach der Definition des Statistischen Bundesamtes

⁶ Diplom an Universitäten und vergleichbare Abschlüsse (Magister, Staatsexamen außer Lehramt, künstlerisches Examen), Diplom an Fachhochschulen, Staatsexamen für ein Lehramt (ohne Differenzierung nach Fachrichtung)

3 Design, Durchführung und Rücklauf

[Erste Befragungswelle] Da die Hochschulen die Kontaktdaten ihrer Absolvent*innen aus Datenschutzgründen nicht veröffentlichen durften, teilten sie dem DZHW lediglich die jeweilige Absolvent*innenanzahl mit. Anschließend sendete das DZHW die passende Anzahl an Erhebungsunterlagen für die erste Befragungswelle postalisch an die jeweiligen Prüfungsämter, die diese an die Zielpersonen weiterverschickten (Adressmittlungsverfahren).

Die Erhebungsunterlagen der ersten Befragungswelle bestanden aus einem Anschreiben (inkl. Datenschutzinformationen), einem Papierfragebogen und einem an das DZHW adressierten portofreien Umschlag zur Rücksendung des ausgefüllten Fragebogens. Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich von Januar 1994 bis vermutlich Mai 1994 – hierbei ist jedoch zu beachten, dass das Ende der Feldzeit nicht genau rekonstruiert werden kann. Angenommen wird der Endzeitpunkt der Erhebung angesichts eines Erinnerungsschreibens des Primärforschungsprojekts mit dem Aufruf zur Teilnahme an der hiesigen Studie im April 1994. Darüber hinaus konnte das DZHW aufgrund des angewendeten Kontaktverfahrens über die Prüfungsämter keinen direkten Einfluss auf den genauen Versandzeitpunkt der Erhebungsunterlagen nehmen. Die dabei erhobene Nettostichprobe der ersten Befragungswelle umfasst 11.314 Fälle, damit ergibt sich eine Rücklaufquote von etwa 40 %.

Um die Erhebungsunterlagen in der zweiten Befragungswelle direkt durch das DZHW verschicken zu können, wurden im Fragebogen der ersten Befragungswelle die Kontaktdaten (Post-Adresse) der Befragungsteilnehmer*innen erfasst. Beim Eingang eines Fragebogens im DZHW wurde sowohl auf den Fragebogen als auch auf den Adressabschnitt des Fragebogens per Paginierstempel eine eindeutige Identifikationsnummer gestempelt und aus allen Adressabschnitten eine Referenzliste von der Identifikationsnummer zur jeweils zugehörigen Adresse erstellt.⁷

[Zweite Befragungswelle] Die Erhebungsunterlagen der zweiten Welle umfassten ein Anschreiben (inkl. Datenschutzinformationen), einen Papierfragebogen, einen Kurzbericht mit Ergebnissen aus der ersten Befragungswelle sowie einen an das DZHW adressierten portofreien Rückumschlag. Die Feldzeit der zweiten Erhebungswelle erstreckte sich von Oktober 1998 bis vermutlich Dezember 1998 (siehe hierzu auch Kapitel 4.2 Episodendaten).

Insgesamt umfasst die Nettostichprobe der zweiten Erhebungswelle 6.737 Absolvent*innen. Bezogen auf die 11.314 Befragten in der ersten Befragungswelle konnte in der zweiten Befragungswelle eine Rücklaufquote von etwa 60 % Prozent erzielt werden.

Eine Übersicht, welche Items der Papierfragebögen in beiden Erhebungswellen vorkommen, ist in Anhang A zu finden.

⁷ Zur Gewährleistung des Datenschutzes wurde der Adressabschnitt vom Fragebogen abgetrennt und die Referenzliste getrennt von den Befragungsdaten auf einem geschützten Server gespeichert.

4 Datenaufbereitung

In dem Nachfolgenden werden die verschiedenen Schritte der Datenaufbereitung näher beschrieben. Der aufbereitete Datensatz wird sowohl im Stata- als auch im SPSS-Format bereitgestellt. Die im Rahmen der Datenedition vorgenommene Anonymisierung wird dabei separat in dem folgenden Kapitel 5 erläutert.

4.1 Personendaten

[Datenübertragung] Die Angaben der Befragten mussten aus den Papierfragebögen zunächst auf Basis eines Codeplans zur weiteren Verarbeitung in ein computerlesbares Format übertragen werden. Zu diesem Zweck wurde auf dem Fragebogen vermerkt, welcher Frage bzw. Teilfrage eine Variable zugeordnet ist, welchen Namen diese Variable trägt und welche numerischen Codierungen für die Antworten der Befragten verwendet werden sollten. Um die Erfassungsreihenfolge festzulegen, wurden die Variablen zusätzlich nummeriert. Im Rahmen der Datenübertragung wurden auch bereits manuelle Konsistenzprüfungen und -korrekturen vorgenommen⁸ und numerische Codierungen für zentrale (halb-)offene Angaben vermerkt.

[Datenstruktur] Die Daten der einzelnen Befragungswellen wurden zunächst zusammengeführt. Die Zuordnung der Fälle erfolgte über die im Rahmen der Feldphase vergebenen Identifikationsnummern der Befragten. Pro befragte Person existiert eine Datenzeile im Datensatz (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen.

[Variablenbenennung] Um die Variablen beider Wellen differenzieren zu können wurden die Variablen der Welle 1 mit dem Präfix „a“ versehen und die der zweiten Welle mit dem Präfix „b“. Darüber hinaus sind Variablen, die im Rahmen der Anonymisierung (vgl. Kapitel 5) neu generiert wurden, mit einem anhand eines Unterstrichs abgetrennten g-Suffix gekennzeichnet (_g#).

[Variablen- und Wertelabelvergabe] Für die Variablen- und Wertelabel wurden Formulierungen des Fragebogens übernommen oder Kurzformen dieser gewählt. Grundlage für die Wertelabel sind je nach Fragetyp die Formulierungen der Antwortoptionen. Die Variablen- und Wertelabels liegen in deutscher Sprache vor. Weiterhin wurden in den Variablen- und Wertelabels keine Schönheitskorrekturen (z.B. Rechtschreibung) vorgenommen. Eine Prüfung auf inhaltliche Korrektheit fand jedoch statt. In allen Variablenlabels ist außerdem ein Kürzel für den Zugangsweg enthalten, auf welchem die jeweilige Variable freigegeben ist (vgl. Datennutzungshinweise und Kapitel 5):

- **r:** Variable ist im Remote-Desktop-SUF und im On-Site-SUF nutzbar.
- **o:** Variable ist nicht 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.

⁸ Die vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebögen dokumentiert und sind daher nicht mehr systematisch rekonstruierbar.

[Codierung fehlender Werte] Fehlende Werte wurden durch negative Werte codiert. Tabelle 2 stellt die für das Absolventenpanel 1993 verwendete Missingsystematik dar. Dabei ist zu beachten, dass keine Unterscheidung zwischen den Missings „keine Angabe“ und „Filter“ vorgenommen wurde, da in den Fragebögen nicht nur explizite Filter (z. B. über „weiter mit“) enthalten sind, sondern auch implizit über die Fragetexte oder -hinweise gefiltert wurde.

Tabelle 2 Systematik für fehlende Werte

Code	Wertelabel
-1	keine Angabe/Filter
-5	keine Teilnahme (Panel)
-9	nicht bestimmbar
-967	anonymisiert

[Besondere Hinweise zu den Tätigkeitstableaus] Der Tätigkeitsverlauf der Befragten wurde über sogenannte Tätigkeitstableaus erfasst (vgl. Frage 4.1 in Welle 1 und Frage 1.1 in Welle 2). Zwischen den Variablen aus diesen Tableaus wurden keine Abgleiche und keine Bereinigungen vorgenommen, sodass hier Inkonsistenzen sowie Überschneidungen existieren können. Darauf hingewiesen sei außerdem, dass unklar ist, was genau die Angabe „läuft noch“ (Code 99 in Welle 1, Code 77 in Welle 2) bei den Endzeitpunkten der einzelnen Tätigkeiten bedeutet, da keine genauen Informationen zum jeweiligen Erhebungszeitpunkt bzw. Fragebogeneingangsdatum vorliegen. Für Fälle, die nur an Welle 1 teilgenommen haben, wurden die Angaben aus dem Tätigkeitstableau wegen zu vieler Unklarheiten zum Missingcode „-1“ recodiert. Insgesamt wird empfohlen, die Variablen aus den Tätigkeitstableaus nur mit Bedacht und Einarbeitung zu analysieren. Die Tätigkeitsverläufe aus den Tätigkeitstableaus wurden für die Fälle, die sich an beiden Wellen beteiligt haben, jedoch zusätzlich auch in einen Episodendatensatz überführt, in dessen Erstellung auch Plausibilisierungsannahmen eingeflossen sind (vgl. Kapitel 4.2).

4.2 Episodendaten

[Datenstruktur] Die Episodendaten enthalten nur die Antworten aus den Tätigkeitstableaus (Frage 4.1 in Welle 1 und Frage 1.1 in Welle 2). Dabei ist eine Episode definiert als ein Zeitraum, in dem eine bestimmte Tätigkeitsart (z. B. Erwerbstätigkeit, Ausbildung) ausgeübt wird bzw. ein konkreter Status (z. B. Elternzeit, Arbeitslosigkeit) besteht. Für jede Episode einer Person existiert jeweils eine Datenzeile (long-Format). Die Episoden wurden fallweise sortiert, das heißt alle Episoden einer Person folgen direkt aufeinander. Verschiedene Tätigkeitsarten im selben Zeitraum wurden jeweils als eigenständige Episode codiert. Wenn Tätigkeiten derselben Art unmittelbar aufeinander folgten oder parallel ausgeübt wurden, wurden sie zu einer Episode zusammengefasst. Daher geht aus den Episodendaten nicht hervor, ob eine Episode eine oder mehrere Tätigkeiten derselben Art umfasst. Das Zusammenführen von Personen- und Episodendaten wird über die Identifikationsnummer der Person (Variable: pid) ermöglicht.

[Tätigkeitsarten] Die in den Tätigkeitstableaus angegebenen Tätigkeiten wurden zu zehn Tätigkeitsarten aggregiert. Das Zuordnungsschema ist in Tabelle 3 dargestellt.

Tabelle 3: Zuordnungsschema Tätigkeitsarten

Tätigkeitsart in den Episodendaten	Tätigkeitsart im Tätigkeitstableau
1 arbeitslos	13 Arbeitslosigkeit 20 Arbeits-/Stellensuche
2 Ausbildung	6 Volontariat, Hospitation 7 Berufliche Ausbildung, Lehre 10 Umschulung 11 Fortbildung 25 Private Weiterbildung
3 Erwerbstätig	1 unselbständig berufstätig 2 selbständig berufstätig 4 Trainee 27 wissenschaftliche Angestellte an Hochschulen 28 wissenschaftliche Angestellte an Forschungseinrichtungen
4 Familienarbeit	15 Hausfrau*-mann 16 Kindererziehung 21 Erziehungsurlaub
5 Jobben	12 Jobben
6 keine Angabe	-1 keine Angabe/Filter für „Verlaufslücken“
7 Studium, Promotion	8 Studium (auch Fernstudium) 9 Promotion
8 Referendariat	5 Referendar*in, Vikar*in, (Anerkennungs)Praktikant*in, Ärzt*in im Praktikum
9 Sonstiges	14 Engagement in Bürgerinitiative, Verein, etc. 17 Urlaub, Reisen 18 Krankheit 19 Wehr-/Zivildienst 23 Wohnungssuche/Umzug 24 Praktikum 26 wissenschaftliche Hilfskraft an Hochschulen 99 Sonstiges, freie Zeit
10 Werk-Honorararbeit	3 Werkarbeit, Honorararbeit 22 Lehrbeauftragte*r

[Plausibilisierungsannahmen] Bei der Erstellung der Episodendaten wurden im Primärforschungsprojekt bestimmte Plausibilisierungsannahmen getroffen:

- Es wurden nur die Tätigkeitsverläufe derjenigen Fälle berücksichtigt, die sich sowohl an der ersten als auch an der zweiten Befragungswelle beteiligt haben.
- Bis zum Dezember 1993 wurden die Tätigkeitsverläufe aus den Angaben der ersten Befragungswelle (Frage 4.1) gebildet.
- Ab Januar 1994 wurden die Tätigkeitsverläufe aus den Angaben der zweiten Befragungswelle (Frage 1.1) gebildet.
- Für Fälle mit noch laufender Tätigkeit in Welle 2 wurde aufgrund des nicht vorliegenden Fragebogeneingangsdatums (vgl. Kapitel 4.1) als ungefähre Angabe der Fragebogenversand herangezogen (Ende Oktober 1998, d. h. die ersten Fragebögen sind im November 1998 zurückgekommen). Als letzter Monat wurde der Dezember 1998 berücksichtigt.

Darauf hinzuweisen ist, dass die Plausibilisierungsannahmen nicht parallel auch in den Variablen aus den Tätigkeitstableaus (vgl. Kapitel 4.1) berücksichtigt wurden, sodass Inkonsistenzen zwischen den Tätigkeitsverläufen in den Episodendaten und in den Tableau-Variablen bestehen können.

5 Anonymisierung

[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten⁹, 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 2017.¹⁰ 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 des DZHW stellt für das Absolventenpanel 1993 ein SUF für die wissenschaftliche Sekundärnutzung 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. Das SUF wird dabei über zwei verschiedene Zugangswege angeboten: Remote-Desktop und On-Site (für weiterführende Informationen vgl. Datennutzungshinweise). Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen in Abhängigkeit vom Zugangsweg erläutert.

[Statistische Anonymisierungsmaßnahmen] Im Rahmen der Anonymisierung sind zunächst alle Informationen, mit denen sich Personen oder Institutionen direkt identifizieren lassen, zu löschen. Diese sogenannten *direkten Identifikatoren*, wie Namen, Adressen und E-Mail Adressen, sind nicht im SUF des Absolventenpanels 1993 enthalten. Zudem wurde 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

⁹ „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).

¹⁰ 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.

indirekt zu identifizieren. Um eine eindeutige Zuordnung der Daten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Zugangsweg – aggregiert oder gar nicht freigegeben. Im Rahmen des Absolventenpanel 1993 sind besonders zentrale Quasi-Identifikatoren beispielsweise die Hochschule und das Studienfach.

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). Die offenen Angaben wurden größtenteils bereits im Rahmen der Datenaufbereitung durch das Primärforschungsprojekt vercodet und werden in dieser Form (teilweise aggregiert) zur Verfügung gestellt. Nicht codierte offene Angaben wurden 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 notwendig 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 (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). Im Absolventenpanel 1993 wurden Gesundheitsinformationen erhoben, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Daher wurden diese Antworten gelöscht.

Die nachfolgende Tabelle 4 stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen je nach Zugangsweg dar. Genauere Informationen zur Aggregation der einzelnen Variablen sind im Datensatzreport zu finden.

Tabelle 4: Maßnahmen der statistischen Anonymisierung nach Zugangsweg

Merkmal	On-Site-SUF	Remote-Desktop-SUF
Direkte Identifikatoren	Generierung einer systemfreien ID	Generierung einer systemfreien ID
Studienfach (offene Angabe) (Welle 1)	Vercodung/Aggregation zu Studienbereichen	Vercodung/Aggregation zu Studienbereichen
Hochschule (offene Angabe) (Welle 1)	Vercodung und Aggregation zu <ul style="list-style-type: none"> - Hochschulart (Universität/Fachhochschule) - Hochschulort (NUTS-2-Region) - Hochschulort (Bundesland) - Hochschulort (Bundesland aggregiert) 	Vercodung und Aggregation zu <ul style="list-style-type: none"> - Hochschulart (Universität/Fachhochschule) - Hochschulort (Bundesland aggregiert)
Region der Beschäftigung (PLZ 3-Steller und Ausland) (Welle 1 und 2)	Freigabe Vercodung und Aggregation zu <ul style="list-style-type: none"> - NUTS-2-Region - Bundesland/Ausland 	Vercodung und Aggregation zu <ul style="list-style-type: none"> - NUTS-2-Region - Bundesland/Ausland
Tätigkeitsart nach Studienabschluss (offene Angabe) (Welle 1 und 2)	Vercodung und Aggregation von Kategorie „Krankheit“ zu „Sonstiges“	Vercodung und Aggregation von Kategorie „Krankheit“ zu „Sonstiges“
Institution der Tätigkeit (offene Angabe) (Welle 1)	Vercodung	Löschung
Art des Stipendiums zur Finanzierung des Studiums/der Promotion (offene Angabe) (Welle 1)	Vercodung und Aggregation zu <ul style="list-style-type: none"> - Forschungsgemeinschaften - Öffentliche Stipendien - Private Stipendien - Sonstiges 	Vercodung und Aggregation zu <ul style="list-style-type: none"> - Forschungsgemeinschaften - Öffentliche Stipendien - Private Stipendien - Sonstiges
Stunden der Teilzeitarbeit (offene Angabe) (Welle 1)	Freigabe	Löschung
Erlerner Ausbildungsberuf (Welle 1)	Vercodung und Aggregation zu <ul style="list-style-type: none"> - KldB-88-2-Steller 	Vercodung und Aggregation zu <ul style="list-style-type: none"> - KldB-88-2-Steller
Berufliche Funktion und Aufgaben (offene Angabe) (Welle 1 und 2)	Löschung	Löschung
Brutto-Monatseinkommen (Welle 1 und 2)	Freigabe	Kategorisierung zu 500er Schritten; über 10.000 DM zusammengefasst
Anzahl der Kinder (Welle 1 und 2)	Aggregation der Kategorien „4“ und „5“ zu „vier Kinder und mehr“	Aggregation der Kategorien „4“ und „5“ zu „vier Kinder und mehr“

Merkmalsname	On-Site-SUF	Remote-Desktop-SUF
Weitere Studienfinanzierung Art der Mittel (offene Angabe) (Welle 1)	Vercodung	Löschung
Gründe für die Erwerbslosigkeit (Welle 2)	Kategorie „aus gesundheitlichen Gründen“ zu „sonstige“	Kategorie „aus gesundheitlichen Gründen“ zu „sonstige“
Genauere Berufsbezeichnung (offene Angabe) (Welle 2)	Vercodung zu - KldB-92-4-Steller	Vercodung und Aggregation zu - KldB-92-3-Steller
Angestrebte Berufsbezeichnung nach Umschulung (offene Angabe) (Welle 2)	Vercodung zu - KldB-92-4-Steller	Löschung
Arbeitsplatzwechsel aus gesundheitlichen Gründen (Welle 2)	Löschung	Löschung
Grund für Beabsichtigung eines Stellenwechsels (Welle 2)	Kategorie „gesundheitliche Gründe“ zu „keine Angabe/Filter“	Kategorie „gesundheitliche Gründe“ zu „keine Angabe/Filter“
Bundesland/Land der Beschäftigung (offene Angabe) (Welle 2)	Vercodung	Vercodung und Aggregation zu - Bundesland (Schleswig-Holstein/Hamburg, Niedersachsen/Bremen, Rheinland-Pfalz/Saarland, restliche Bundesländer werden nicht aggregiert)/Ausland
Kindesalter (Welle 2)	Freigabe	Löschung
Sonstige offene Nennungen (Welle 1 und 2)	Vercodung oder Löschung	Vercodung oder Löschung

6 Literaturverzeichnis

- Ebel, T. (2015). *Empfehlungen zur Anonymisierung quantitativer Daten*. Mannheim: GESIS - Leibniz-Institut für Sozialwissenschaften.
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). *Datenschutz am Forschungsdatenzentrum (FDZ-Methodenreport Nr. 6)*. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Holtkamp, R., Minks, K.-H. & Schaeper, H. *Fachhochschulabsolventen im Strukturwandel des Beschäftigungssystems*. Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie: Bonn.
- 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.
- Rehn, T., Brandt, G., Fabian, G., & Briedis, K. (2011). Hochschulabschlüsse im Umbruch. (Forum Hochschule Nr. F17/2011) *Studium und Übergang von Absolventinnen und Absolventen reformierter und traditioneller Studiengänge des Jahrgangs 2009* (HIS:Forum Hochschule 17/2011). Hannover: HIS.
- Schramm, M. & Beck, S. (2010). *Dokumentation des Scientific Use Files „HIS Absolventenpanel 2001“*. Datensatzbeschreibung und Kodebuch. Hannover: HIS GmbH.

Anhang

A. Referenzliste: Fragen, die in beiden Befragungswellen vorkommen

Welle 1	Welle 2
Frage 1.19	Frage 1.6
Frage 2.2	Frage 3.1
Frage 2.3	Frage 3.3
Frage 2.5	Frage 3.2*
Frage 4.1	Frage 1.1*
Frage 4.2	Frage 9.1
Frage 4.12	Frage 9.2
Frage 4.15	Frage 1.4
Frage 4.17	Frage 6.1*
Frage 5.1	Frage 8.1
Frage 5.3	Frage 6.3*
Frage 5.4	Frage 8.2
Frage 5.5	Frage 8.8
Frage 5.6	Frage 8.13 und 8.14
Frage 5.7	Frage 8.5
Frage 5.8	Frage 8.3
Frage 5.9	Frage 8.15 und 8.16
Frage 5.10	Frage 8.4
Frage 5.11	Frage 8.7*
Frage 5.12	Frage 8.18
Frage 5.13	Frage 1.3
Frage 6.10	Frage 9.8
Frage 6.11	Frage 9.9
Frage 6.12	Frage 9.10

* Fragen nur ähnlich, nicht identisch