

Anett Friedrich

# BIBB/BAuA-Erwerbstätigen- befragung 2012 – Ausbildungs- absolventen der Jahre 2006 bis 2008

Version 1.1



Heft 2/2016

Anett Friedrich

# **BIBB/BAuA-Erwerbstätigen- befragung 2012 – Ausbildungs- absolventen der Jahre 2006 bis 2008**

Version 1.1

### ► Danksagung

Wir danken dem Projektmitarbeiter des Teilprojekts der Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008, Ralf Dorau, für seine wertvolle Hilfe bei der Erstellung des Handbuchs.

### ► Datenverfügbarkeit

Der in diesem Handbuch beschriebene Datensatz kann als Scientific-Use-File über unsere Internetseite [www.bibb-fdz.de](http://www.bibb-fdz.de) beantragt werden.

### ► Hinweis zur Zählweise bei Versionsnummern des Datensatzes

Änderungen gegenüber der Vorversion ohne größere inhaltliche Relevanz werden durch fortlaufende Nummern *nach* dem Punkt dokumentiert (zweite Ebene). Inhaltlich relevante Änderungen führen demgegenüber zu einer fortlaufenden Nummerierung auf der ersten Ebene.

## Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

© 2016 by Bundesinstitut für Berufsbildung, Bonn

BIBB-FDZ Daten und Methodenbericht 2/2016, Version 1.0

### Herausgeber:

Bundesinstitut für Berufsbildung, Bonn  
Internet: [www.bibb.de](http://www.bibb.de)

### Publikationsmanagement:

AB 1. 4

### Gesamtherstellung:

W. Bertelsmann Verlag GmbH & Co. KG

### Vertriebsadresse:

Bundesinstitut für Berufsbildung  
Arbeitsbereich 1.4 – Publikationsmanagement/Bibliothek  
– Veröffentlichungen –  
Robert-Schuman-Platz 3  
53175 Bonn  
E-Mail: [zentrale@bibb.de](mailto:zentrale@bibb.de)

### Kontakt FDZ:

E-Mail: [fdz@bibb.de](mailto:fdz@bibb.de)  
Tel.: +49 228 107-2041  
Fax: +49 228 107-2020

Printed in Germany

ISSN: 2190-300X



Der Inhalt dieses Werkes steht unter einer Creative-Commons-Lizenz (Lizenztyp: Namensnennung – Keine kommerzielle Nutzung – Keine Bearbeitung – 4.0 Deutschland).

Weitere Informationen finden Sie im Internet auf unserer Creative-Commons-Infoseite [www.bibb.de/cc-lizenz](http://www.bibb.de/cc-lizenz).

Diese Netzpublikation wurde bei der Deutschen Nationalbibliothek angemeldet und archiviert: [urn:nbn:de:](http://nbn-resolving.org/urn:nbn:de:dnb:1-107-2016-0001-0)

Internet: [www.bibb.de/veroeffentlichungen](http://www.bibb.de/veroeffentlichungen)

Downloads unter: [www.bibb-fdz.de](http://www.bibb-fdz.de)

# Inhaltsverzeichnis

<b>Tabellen</b> .....	5
<b>1 Einleitung</b> .....	6
1.1 Gegenstand des Handbuchs .....	6
1.2 Aufbau des Handbuchs .....	6
<b>2 Die BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 im Überblick</b> .....	7
2.1 Kurzbeschreibung und Aufbau .....	7
2.2 Zentrale Eigenschaften .....	8
<b>3 Inhalte der Erhebung</b> .....	9
<b>4 Datenerhebung</b> .....	10
<b>5 Datenaufbereitung</b> .....	13
5.1 Querschnittdatensatz .....	13
5.2 Ereignisdatensatz .....	14
5.3 Klassifikationen .....	15
5.4 Textvariablen .....	15
<b>6 Datenzugang und Datenanonymisierung</b> .....	17
6.1 Datenzugang .....	17
6.2 Bestimmungen des Datenschutzes .....	17
6.3 Anonymisierungen nach Datenprodukten .....	18
6.3.1 Anonymisierung: SUF .....	18
6.3.2 Formale Anonymisierung: Datenfernverarbeitung .....	19
6.3.3 Projektbezogene Anonymisierung: Gastrechnen im BIBB .....	19
<b>7 Arbeitshilfen</b> .....	21
7.1 Gewichtung .....	21
7.2 Deklarierung von Filterwerten und fehlenden Werten nach dem BIBB-FDZ-Standard .....	21
7.3 Zitierhinweise .....	22
<b>Literatur</b> .....	23

## Tabellen

<b>Tab. 1:</b>	Fallzahlen nach Rekrutierungswegen . . . . .	10
<b>Tab. 2:</b>	Häufigkeiten der erfassten Aktivitäten . . . . .	11
<b>Tab. 3:</b>	Anzahl der Spells nach Aktivitätsart . . . . .	11
<b>Tab. 4:</b>	Häufigkeiten der erfassten Umzüge . . . . .	12
<b>Tab. 5:</b>	Übersicht der gebildeten Klassifikationsvariablen . . . . .	15
<b>Tab. 6:</b>	Textvariablen der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 . . . . .	15
<b>Tab. 7:</b>	Anonymisierte Originalvariablen im SUF der BIBB/BAuA-Erwerbstätigen- befragung 2011/2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 . . . . .	18
<b>Tab. 8:</b>	Vergrößerte bzw. nicht verfügbare Variablen für die DFV der BIBB/BAuA- Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 . . . . .	19
<b>Tab. 9:</b>	Anonymisierte Variablen im GWA der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 . . . . .	20
<b>Tab. 10:</b>	Standarddeklaration des BIBB-FDZ für verschiedene Arten fehlender Antwortwerte . . . . .	21

# 1 Einleitung

## 1.1 Gegenstand des Handbuchs

Der vorliegende Daten- und Methodenbericht beschreibt detailliert die Forschungsdaten der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008.

Die Daten wurden von TNS Infratest Sozialforschung, München (TNS Infratest), im Auftrag des Bundesinstituts für Berufsbildung (BIBB) erhoben. Das Forschungsdatenzentrum im BIBB (BIBB-FDZ) hat weitere Aufbereitungsschritte an den Daten vorgenommen, die im vorliegenden Handbuch dokumentiert sind. Für die externe nicht kommerzielle Forschung bestehen Datenzugangsmöglichkeiten zu den Forschungsdaten der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 in Form von Scientific-Use-Files, über die Datenfernverarbeitung oder einen Gastaufenthalt.

## 1.2 Aufbau des Handbuchs

Das nächste Kapitel gibt einen Überblick über die zentralen Merkmale der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008. Kapitel 3 skizziert die Inhalte der Befragung. Kapitel 4 widmet sich der Datenerhebung, bevor im fünften Kapitel die Aufbereitungsschritte für die Forschungsdaten beschrieben werden. Anschließend wird in Kapitel 6 aufgezeigt, in welcher Weise die Anonymisierungen der Daten für die jeweiligen Datenzugangswege vorgenommen wurden. Kapitel 7 erläutert abschließend die Arbeitshilfen der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008.

## 2 Die BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 im Überblick

### 2.1 Kurzbeschreibung und Aufbau

Bei der Erhebung von Ausbildungsabsolventen der Jahre 2006 bis 2008 handelt es sich um eine Ergänzungsstichprobe der BIBB/BAuA-Erwerbstätigenbefragung 2011/2012. Sie ist Teil eines Gesamtprojekts, welches aus insgesamt drei Befragungen besteht: der Haupterhebung der BIBB/BAuA-Erwerbstätigenbefragung<sup>1</sup> (Los 1), der Jugenderwerbstätigenbefragung 2012<sup>2</sup> (Los 2) und eben der Befragung von Ausbildungsabsolventen der Jahre 2006 bis 2008 (Los 3).

Zur Zielgruppe der Los-3-Erhebung gehören alle Personen, die zwischen 2006 und 2008 ihre berufliche Erstausbildung in Deutschland abgeschlossen haben. Die BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 wurde, teilweise zeitlich parallel mit den anderen beiden Erhebungen, zwischen Oktober 2011 und Dezember 2012 durchgeführt. Die Datenerhebung erfolgte mittels drei verschiedener Rekrutierungswege: über Los 1 und Los 2, über eine Mehrthemenbefragung und über Online-Rekrutierung.

Das BIBB-FDZ bietet mit einem Querschnitt- und einem Ereignisdatensatz zwei separate Forschungsdatensätze der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 an. Der Querschnittdatensatz enthält vor allem die Informationen, die über die einzelnen erfassten (Aus-)Bildungs- und Berufsetappen konstant sind (u. a. Soziodemografie und Informationen zur ersten beruflichen Ausbildung). Der Ereignisdatensatz enthält für jeden der 4.711 Befragten im Mittel zwei biografische Episoden zum weiteren Bildungs- und Berufsverlauf nach der Erstausbildung sowie mindestens eine Umzugsepisode. Insgesamt enthält der Ereignisdatensatz 16.897 verschiedene Spells, aufgebaut ist er im sogenannten „langen Format“, d. h., die Längsschnittinformationen sind zeilenweise abgelegt. Neben einer Personen-ID (*Intnr*, systemfreies Kennzeichen) enthält der Datensatz für jede Person mindestens zwei Spell-IDs (*Spell*) und die zu jeder erfassten Aktivität erhobenen Variablen.

<sup>1</sup> Nähere Infos gibt es unter <https://metadaten.bibb.de/metadaten/27>.

<sup>2</sup> Nähere Infos gibt es unter <https://metadaten.bibb.de/metadaten/77>.



## 2.2 Zentrale Eigenschaften

Befragungstitel	BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008
DOI	10.7803/401.12.3.1.10
Kurzbeschreibung	Die BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 beinhaltet Daten zum Übergang von der Ausbildung in die Erwerbstätigkeit, also zur sogenannten zweiten Schwelle. Erfasst wird ein Zeitraum von drei Jahren, um sowohl erfolgreiche Berufseinstiegsprozesse sowie prekäre Verläufe zu erfassen.
Erhebungsjahr	2011/2012
Erhebungseinheit	Ausbildungsabsolventen
Themenschwerpunkt	Zweite Schwelle
Datenzugangsmöglichkeiten	Scientific-Use-File, Gastwissenschaftleraufenthalt, Datenfernverarbeitung
Variablenanzahl	97 (Querschnittdatensatz); 92 (Ereignisdatensatz)
Grundgesamtheit	Personen, die in den Jahren 2006 bis 2008 ihre erste berufliche Ausbildung in Deutschland abgeschlossen
Gewichtung/Hochrechnung	Populationsgewicht (West/Ost, Geschlecht, Wirtschaftsbereich; Geschlecht, Uni/FH)
repräsentative Region	West/Ost
Fallzahl	4.711 Personen und 16.897 Ereignisse (davon 5.935 Umzüge)
Erhebungsverfahren	CATI, Mehrthemenbefragung, Online-Rekrutierung
Auswahlverfahren	ADM (Weiterentwicklung des Gabler-Häder-Verfahrens), Kish-Methode
Erhebungsdesign	Querschnitt
Bemerkung	Die Erhebung ist Teil eines Gesamtprojekts welches aus insgesamt drei Befragungen besteht: der Haupterhebung der BIBB/BAuA Erwerbstätigenbefragung 2012 (Los 1), der Jugenderwerbstätigenbefragung 2012 (Los 2) und eben der Befragung von Ausbildungsabsolventen der Jahre 2006 bis 2008 (Los 3).
Links	BIBB-FDZ Metadatenportal: <a href="http://metadaten.bibb.de/metadaten/87">http://metadaten.bibb.de/metadaten/87</a>
Stichworte	Berufseinstieg, Bildungswege, Ausbildung, berufliche Integration, Berufseinstiegsprozesse, Übergänge von Ausbildung in Erwerbstätigkeit

## 3 Inhalte der Erhebung

Der thematische Schwerpunkt der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 ist die sogenannte zweite Schwelle, also der Übergang von der Ausbildung in die Erwerbstätigkeit. Erfasst wird dabei ein Zeitraum von drei Jahren, um erfolgreiche Berufseinstiegsprozesse von potenziell prekären Verläufen und Entwicklungen abgrenzen zu können.

Der Verbleib der Ausbildungsabsolventen wurde detailliert erfasst, unterschieden wurden folgende Aktivitäten:

1. Erwerbstätigkeit (Frageblock A2100)
2. Arbeitslosigkeit (Frageblock A2200)
3. Phasen von schulischer oder beruflicher Weiterbildung (Frageblöcke A2400, A2500, A2600 und A2700)
4. Etwas anderes (Frageblöcke A2800 und A2900).

Zu den einzelnen Aktivitäten wurden unterschiedlich viele Fragen gestellt. So umfasst etwa die Phase der Erwerbstätigkeit bis zu 21 Fragen (etwa zur beruflichen Stellung, zum monatlichen Bruttolohn [in Kategorien] oder zur Befristung), wohingegen zur Phase der Arbeitslosigkeit nur zwei Fragen gestellt wurden (wurde Arbeitslosengeld oder Arbeitslosenhilfe bezogen und Ende der Arbeitslosigkeit).

Weiterhin wurde die Umzugsbiografie (Frageblock A5000) der Befragten erfasst. Diese enthält bis zu acht Umzüge, deren Zeitpunkt, das Bundesland des neuen Wohnorts und ob der bisherige Wohnort mehr oder weniger als 50 Kilometer vom neuen entfernt liegt.

Zusätzlich zu den einzelnen Episoden wurden Merkmale der ersten Berufsausbildung (bspw. Art der Ausbildung, genaue Berufsbezeichnung sowie die Abschlussnote) und soziodemografische Merkmale (etwa Schulabschluss, soziale und ethnische Herkunft) erhoben. Darüber hinaus sollten die Jugendlichen einschätzen, für wie gelungen sie den Einstieg ins Berufsleben halten (Frage A4001).

## 4 Datenerhebung

Die Datenerhebung erfolgte durch TNS Infratest, Details zu selbiger können dem Feldbericht von GENSICKE/TSCHERSICH 2012 entnommen werden. Ein kurzer Überblick wird im Folgenden gegeben.

Die Grundgesamtheit der Erhebung sind Personen, die in den Jahren 2006 bis 2008 ihre erste berufliche Ausbildung in Deutschland abgeschlossen haben und telefonisch zu erreichen waren. Die Auswahl der Zielpersonen erfolgte über die Los-1- und Los-2-Erhebungen<sup>3</sup>, über eine Mehrthemenbefragung und via Online-Rekrutierung. Letztere Auswahlmöglichkeit wurde eingesetzt, da sich das Erreichen der Zielfallzahl ohne diesen zusätzlichen Rekrutierungsweg als problematisch erwiesen hätte.

Die Auswahl der Haushalte erfolgte über das Gabler-Häder-Verfahren und die Auswahl der zu befragenden Person über die Kish-Methode. Die Interviews wurden alle als Computer Assisted Telephone Interviews (CATI)<sup>4</sup> durchgeführt.

Der Erhebungszeitraum lag zwischen November 2011 und Dezember 2012. Die Fallzahlen getrennt nach Rekrutierungswegen können Tabelle 1 entnommen werden. Die durchschnittliche Interviewdauer betrug 16 Minuten.

**Tabelle 1**

Fallzahlen nach Rekrutierungswegen

<i>Rekrutierungsweg</i>	<i>Fallzahl</i>
Los-1-Erhebung	517
Los-2-Erhebung	62
Mehrthemenbefragung	3.064
Online-Rekrutierung	589

Für die Befragten wurden bis zu 16 Aktivitäten bzw. Spells zum Bildungs- und Berufsverlauf erfasst (vgl. Tabelle 2). Da der Fragebogen bis zu 20 Schleifen für die Erfassung dieser vorsah, ist davon auszugehen, dass der Verbleib der Befragten nach der Erstausbildung vollständig erhoben wurde. Etwa 85 Prozent der Befragten haben Angaben für bis zu drei Aktivitäten gemacht. Insgesamt enthält der Ereignisdatensatz 10.962 Spells zum Bildungs- und Berufsverlauf, deren Verteilung auf die verschiedenen Aktivitätsarten Tabelle 3 entnommen werden kann. Über die Hälfte der erfassten Aktivitäten befassen sich mit der Erwerbstätigkeit, zu welcher auch die meisten Fragen gestellt wurden. Der Anteil der anderen drei Aktivitäten liegt zwischen knapp 11 und ca. 17 Prozent.

<sup>3</sup> Die Los-1- und Los-2-Erhebungen wurden als Screening genutzt, es wurde ermittelt, ob der Befragte zwischen 2006 und 2008 die erste berufliche Ausbildung abgeschlossen hat. War dies nicht der Fall, wurde gefragt, ob eine andere im Haushalt lebende Person im entsprechenden Zeitraum eine berufliche Erstausbildung abgeschlossen hat.

<sup>4</sup> Die Personen, die via Online-Access-Panel rekrutiert wurden, wurden nach ihrer Bereitschaft gefragt, ein Telefoninterview zu geben, und anschließend nach einer entsprechenden Telefonnummer.

**Tabelle 2**  
Häufigkeiten der erfassten Aktivitäten

<i>Anzahl Spells</i>	<i>Häufigkeit</i>	<i>Prozent</i>	<i>Kumulierte Prozent</i>
1	4.711	42,98	42,98
2	3.078	28,08	71,05
3	1.595	14,55	85,60
4	758	6,91	92,52
5	369	3,37	95,89
6	194	1,77	97,66
7	113	1,03	98,69
8	59	0,54	99,22
9	33	0,30	99,53
10	20	0,18	99,71
11	12	0,11	99,82
12	10	0,09	99,91
13	6	0,05	99,96
14	2	0,02	99,98
15	1	0,01	99,99
16	1	0,01	100,00
Summe	10.962	100,00	100,00

**Tabelle 3**  
Anzahl der Spells nach Aktivitätsart

<i>Aktivitätsart</i>	<i>Häufigkeit</i>	<i>Prozent</i>	<i>Kumulierte Prozent</i>
1. erwerbstätig	6.450	58,84	58,84
2. arbeitslos	1.530	13,96	72,80
3. schulische oder berufliche Weiterbildung	1.820	16,60	89,40
4. etwas anderes	1.152	10,51	99,91
9. keine Angabe	10	0,09	100,00
Summe	10.962	100,00	100,00

Neben den Spells des Bildungs- und Berufsverlaufs nach der Erstausbildung wurden 5.935 Umzugsspell erfasst (vgl. Tabelle 4). Jeder der Befragten weist mindestens einen Umzugsspell auf, da gefragt wurde, ob nach der Erstausbildung ein Wohnortwechsel stattfand oder nicht, und diese Information im ersten Umzugsspell enthalten ist. Etwa 20 Prozent der Befragten weisen mehr als einen Umzug auf.

**Tabelle 4**  
Häufigkeiten der erfassten Umzüge

<i>Anzahl Umzugsspells</i>	<i>Häufigkeit</i>	<i>Prozent</i>	<i>Kumulierte Prozent</i>
1	4.711	79,38	79,38
2	870	14,66	94,04
3	252	4,25	98,28
4	67	1,13	99,41
5	23	0,39	99,80
6	10	0,17	99,97
7	1	0,02	99,98
8	1	0,02	100,00
Summe	5.935	100,00	100,00

## 5 Datenaufbereitung

Durch die Entwicklung von Filterführung und Prüfmodulen durch die Projektmitarbeiter/-innen im BIBB und deren programmiertechnische Umsetzung in ein CATI-Programm durch TNS Infratest ist die Qualität der Daten von vornherein als gut einzuschätzen. Vom Umfrageinstitut wurde dem BIBB ein Datensatz übergeben, welcher anschließend in einen Querschnitt- und einen Ereignisteil geschnitten wurde. Zuvor wurden allerdings aus dem gelieferten Datensatz 61 Fälle entfernt. Dies geschah aus folgenden Gründen:

- ▶ 6 Fälle haben bei der Frage A1012 das Endjahr der Ausbildung nicht angegeben,
- ▶ 14 Fälle haben angegeben, die Ausbildung mit unter 14 Jahren begonnen zu haben,
- ▶ 4 Fälle waren doppelt im Datensatz enthalten,
- ▶ 37 Fälle haben die Ausbildung im Ausland beendet und gehören somit nicht zur Grundgesamtheit.

Schließlich wurden im BIBB-FDZ noch weitere Datenaufbereitungen durchgeführt. Hierzu gehören: die Prüfung und ggf. Korrektur der Filterführung, das Hinzufügen von Hilfs- sowie Klassifikationsvariablen (siehe Abschnitt 5.1, 5.2 und 5.3), die Anonymisierungen (Kapitel 6) sowie die Standarddeklarierung der fehlenden Werte (Abschnitt 7.2).

Über das BIBB-FDZ sind vier Forschungsdatensätze der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 verfügbar: ein Scientific-Use-File, bestehend aus einem Querschnitt- (vgl. Abschnitt 5.1) und einem Ereignisteil (vgl. Abschnitt 5.2), ein Datensatz mit den Textvariablen, ein Datensatz mit weiteren Sondervariablen (vgl. Abschnitt 5.3) sowie ein Forschungsdatensatz mit regionalen Identifikatoren.

### 5.1 Querschnittdatensatz

Der Querschnittdatensatz enthält die spellinvarianten Originalvariablen. Wenn nicht explizit erwähnt, sind die direkt erhobenen Variablen im Datensatz nach der Fragennummer im Fragebogen benannt. Neben den direkt erhobenen Variablen enthält der Datensatz ebenfalls geänderte bzw. nachträglich gebildete Variablen.

Auf Basis der Spell-Identifikationsnummer des Ereignisdatensatzes wurde die Variable Anzahl der Spells pro Befragter (*Nspell*) gebildet und dem Querschnittdatensatz zugefügt. Wobei die Umzugsspellings nicht in die Zählung eingegangen sind.

Aus den Angaben zum Interviewjahr und -monat sowie zum Geburtsjahr und -monat wurde für jeden Befragten, sofern die Angaben zum Geburtsdatum vollständig waren<sup>5</sup>, das Alter zum Befragungszeitpunkt berechnet. Im Durchschnitt liegt das Alter der Befragten bei 27 Jahren, wobei die Spanne von 19 bis 64 Jahren reicht.

Die Variablen *A1003* und *A1003a* waren im dem BIBB-FDZ übergebenen Datensatz nicht enthalten. Anhand der Angaben aus den Variablen *AUS2006*, *A1001* und *A1002* wurden die beiden Variablen allerdings vom BIBB-FDZ nachgebildet.

Zwei Variablen, welche laut Fragebogen vom Erhebungsinstitut erhoben bzw. berechnet wurden, sind nicht im Datensatz enthalten. Es handelt sich um die Variablen *P3005* und

<sup>5</sup> In 8 Fällen waren die Angaben nicht vollständig und wurden entsprechend mit -9 "keine Angabe" codiert.

*P3005kor\_A12014*. Beide Variablen sind Prüfvariablen, die testen sollen, ob der Schulabschluss nach der Erstausbildung höher ist als der erste erworbene Schulabschluss (*P3005*). Traf diese Bedingung nicht zu, wurde der Schulabschluss ggf. in der Variable *P3005kor\_A12014* korrigiert. Die Variable *P3005* ist in 17 Fällen belegt, davon 16-mal mit dem Code 3, der für „keine Korrektur“ steht und einmal mit Code 1 „Abschluss nach erster Berufsausbildung korrigieren“. Diese eine Korrektur wurde während des Interviews in den Daten vorgenommen.

Bei den Variablen *A3004\_em* und *A3004\_ej* wurde jeweils eine Kategorie gebildet, um Fälle abgrenzen zu können, bei denen der abgefragte Schulbesuch noch andauert. Bei allen Fällen, welche bei Variable *A3002* den Wert 4 (Nein, bisher kein Abschluss, bin noch in Schule/Ausbildung/Weiterbildung/Studium) oder bei Variable *A3003* den Wert 3 (Bin noch in Schule/Ausbildung/Weiterbildung/Studium) aufwiesen, wurden *A3004\_em* und *A3004\_ej* von einem Systemmissing auf die Werte 98. bzw. 9998. „Ausbildung noch nicht abgeschlossen“ umgesetzt.

## 5.2 Ereignisdatensatz

Der Ereignisdatensatz enthält alle direkt erhobenen Originalvariablen des Längsschnittteils. Wenn nicht explizit erwähnt, sind die direkt erhobenen Variablen im Datensatz nach der Fragennummer im Fragebogen benannt. Neben den direkt erhobenen enthält der Datensatz nachträglich gebildete oder geänderte Variablen, welche im Folgenden erläutert werden.

Zunächst enthält der Datensatz zwei synthetische Variablen, die die Verschränkung der Quer- und Längsschnittinformationen organisieren. Die Variable *Intnr* ist die originale systemfreie Personenkennziffer und *Spell* die generierte Spell-Identifikationsnummer. Letztere enthält zum einen die Identifikationsnummer des Bildungs- und Berufsverlaufs (zwischen 1 und 16) und zum anderen die der Umzugsbiografie (zwischen 101 und 108).

Für die kalendarische Verortung eines jeden biografischen Spells wurden die spellspezifischen Informationen über das Ende (jeweils Monat und Jahr) einer jeden Aktivität in zusätzliche, übergreifende Kalendervariablen abgelegt: *Endmo* und *Endja*. Diese Variablen unterscheiden sich von den jeweiligen Originalvariablen insofern, dass den Jahreszeitangaben entsprechende Monate zugewiesen<sup>6</sup> wurden. Da bei den einzelnen Episoden immer nur das Ende abgefragt wurde, enthalten die Variablen *Begmo* und *Begja* jeweils die Informationen des Endjahres bzw. Endmonats des vorherigen Spells plus einen Monat. Beim ersten Spell wird das Enddatum der ersten abgeschlossenen Berufsausbildung plus einen Monat als Anfangsdatum codiert. Aufgrund dieses Vorgehens der Berechnung der Beginndaten eines Spells sind zeitlich parallel auftretende Spells nicht möglich. Dies wird jedoch in Kauf genommen, da im Fragebogen explizit nach der sich anschließenden Aktivität gefragt wurde. Weiterhin werden mögliche Lücken in der Erfassung der Episoden nicht berücksichtigt, dies muss bei der Berechnung von Übergangsdauern beachtet werden, da diese sonst ggf. als zu kurz eingeschätzt werden.

Die laut Fragebogen generierte Variable *dauer*, welche im Zuge des Interviews als Kontrollvariable automatisch gebildet wurde, war als solche nicht im an das BIBB-FDZ übergebenen Datensatz enthalten. Eine neue Variable *dauer* wurde vom BIBB-FDZ mithilfe folgender Variablen berechnet: anhand der Angaben der Variablen zu Frage A2113 und dem Endmonat bzw. Endjahr des vorherigen Spells, wenn die Erwerbstätigkeit beendet ist, oder dem Interviewdatum und dem Endmonat bzw. Endjahr des vorherigen Spells, wenn die Erwerbstätigkeit noch andauert.

<sup>6</sup> Die Regel zur Zuordnung der ungefähren Monatsangaben lautet: 21 (Jahresanfang) → 1; 27 (Sommer, Jahresmitte) → 7; 30 (Herbst) → 10; 31 (Winter/Jahresende) → 12 (Dezember) und 99 (keine Angabe) → 6, falls eine gültige Jahresvariable vorlag.

Weiterhin nicht im an das BIBB-FDZ übergebenen Datensatz enthalten war die Kontrollvariable P2109. Auch diese wurde für den SUF nachgebildet. Genutzt wurden dafür die Variablen A2100, A2109A und A2109A. Nicht im Datensatz enthalten ist dagegen die Prüfvariable P1212.

### 5.3 Klassifikationen

Mithilfe der Standardprogramme aus dem BIBB-FDZ Metadatenportal wurden folgende Klassifikationsvariablen<sup>7</sup> gebildet (siehe Tabelle 5). Die entsprechenden Programme können im Metadatenportal heruntergeladen werden. Die entsprechenden Variablen sind im Querschnitt- und Ereignisdatsatz integriert und somit über den SUF nutzbar.

**Tabelle 5**  
Übersicht der gebildeten Klassifikationsvariablen

Variablenname	Variableninhalt	Datensatz
Ta1005_berufsabschnitte	Berufsabschnitte für den 1. Ausbildungsberuf laut KldB 1992	Q
Ta1005_pr_dl	Einteilung in Produktion/Dienstleistung für den 1. Ausbildungsberuf	Q
Ta2101_berufsabschnitte	Berufsabschnitte für Erwerbsberuf laut KldB 1992	E
Ta2101_berufsfeld	Berufsfelddefinition des BIBB aus Erwerbsberuf laut KldB 1992	E
Ta2702_berufsabschnitte	Berufsabschnitte für Studium laut KldB 1992	E
Ta3006_berufsabschnitte	Berufsabschnitte für den 2. Ausbildungsberuf laut KldB 1992	Q
Ta3006_pr_dl	Einteilung in Produktion/Dienstleistung für den 2. Ausbildungsberuf	Q
Mig	Migrationshintergrund	Q

### 5.4 Textvariablen

Zusätzlich zu den numerischen Variablen im Querschnitt- und Ereignisdatsatz stehen die Textvariablen der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 für Analysen zur Verfügung. Aus Datenschutzgründen bzw. aufgrund eines höheren Reidentifikationsrisikos können die Textvariablen nicht über den SUF analysiert werden. Die Textvariablen des Querschnitt- bzw. des Ereignisdatenteils sind als gemeinsamer Datensatz verfügbar und enthalten die der Tabelle 6 zu entnehmenden Textvariablen.

**Tabelle 6**  
Textvariablen der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008

Fragebogensnr.	Variableninhalt	Variablenname
1013	Sonstige Gründe für kurze Ausbildungsdauer (< 12 Monate)	TA1013
1005	Fach- oder Berufsbezeichnung der Erstausbildung	TA1005
1006	Land, in dem der Abschluss erworben wurde	TA1006

<sup>7</sup> Die Bildung weiterer Klassifikationsvariablen ist aufgrund von geringen Fallzahlen ausgeschlossen.



*(Fortsetzung Tab. 6)*

Frage- bogen- nr.	Variableninhalt	Variablenname
1009	Bezeichnung der Bildungseinrichtung der Erstausbildung	TA1009
1011	Ausbildungsstätte der Erstausbildung	TA1011
2101	Tätigkeitsbezeichnung der Erwerbstätigkeit (alle Schleifen)	TA2101_*
2102	Genauere Tätigkeitsbezeichnung der Erwerbstätigkeit (alle Schleifen)	TA2102_*
2503	Fach- oder Berufsbezeichnung der Ausbildung nach Erstausbildung (alle Schleifen)	TA2503_*
2603	Fach- oder Berufsbezeichnung der beruflichen Weiterbildung nach Erstausbildung (alle Schleifen)	TA2603_*
2702	Fach- oder Berufsbezeichnung des Studiums nach Erstausbildung (alle Schleifen)	TA2702_*
5002	In welchem Land liegt der neue Wohnort (alle Schleifen)	TA5002_*

## 6 Datenzugang und Datenanonymisierung

### 6.1 Datenzugang

Der Querschnitt- und der Ereignisdatensatz zur BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 können als Scientific-Use-File (SUF) über die Datenfernverarbeitung und an den Gastarbeitsplätzen im BIBB in Bonn analysiert werden. Diese drei Formen des Zugangs unterscheiden sich sowohl hinsichtlich des Anonymisierungsgrades der Forschungsdaten als auch in der Art der Datenbereitstellung.

Scientific-Use-Files bedienen Forschungsinteressen und werden den Nutzern per Download zur Verfügung gestellt. Dies setzt eine anonymisierte Aufbereitung der Daten voraus. Die Nutzung dieser Forschungsdaten ist an Richtlinien gebunden und erfordert eine förmliche Genehmigung durch das BIBB-FDZ.

Die kontrollierte Datenfernverarbeitung erlaubt die Analyse formal anonymisierter Originaldaten über die Verarbeitung lauffähiger, nutzererstellter Syntaxprogramme im BIBB-FDZ. Die Datennutzer/-innen können hierfür von den Internetseiten des BIBB-FDZ Strukturdatensätze (Spieldaten) herunterladen<sup>8</sup>, die in Aufbau und Merkmalsausprägungen dem Originalmaterial gleichen. Mittels dieser Spieldaten können Auswertungsprogramme in den Analyseprogrammen Stata oder SPSS erstellt werden, mit denen das BIBB-FDZ anschließend die Originaldaten auswertet. Der auf Vertraulichkeit geprüfte Output wird schließlich an die Nutzer/-innen zurückgeschickt.

Gastwissenschaftleraufenthalte erlauben die Analyse schwach anonymisierter Mikrodaten an den abgeschotteten PC-Arbeitsstationen des BIBB-FDZ in einem eigenen Gästeraum. Über einen Gastaufenthalt lassen sich beispielsweise die aus den SUFs entfernten Sondervariablen bzw. vergrößerten Merkmale in ihrem originalen Wertebereich analysieren.

Die Nutzung der Daten erfordert für alle Datenzugangswege eine förmliche Beantragung. Die Nutzungsanträge für alle Zugänge können von den Internetseiten des BIBB-FDZ heruntergeladen werden.

### 6.2 Bestimmungen des Datenschutzes

Der Zugang zu den SUF-Daten des BIBB-FDZ erfolgt ausschließlich unter dem Gebot der Datensparsamkeit und der Einhaltung geltender Datenschutzregeln, insbesondere Artikel 89 der Verordnung (EU) 2016/679 (Datenschutz-Grundverordnung DS-GVO) und § 27 Bundesdatenschutzgesetz (BDSG 2018). Demnach dürfen für die Durchführung wissenschaftlicher Vorhaben Forschungsdaten zeitlich befristet an die unabhängige wissenschaftliche Forschung weitergegeben werden, wenn die Herstellung eines Bezugs zu einer Erhebungseinheit nicht möglich ist ("Anonymität"). Um dies ausnahmslos herzustellen, sind vom Daten empfangenden Institut bzw. den Forschenden besondere technische und organisatorische Datensicherungsmaßnahmen einzurichten, die Unbefugten keinen Datenzugriff ermöglichen. Es wird empfohlen, dass die zugangs- und nutzungsberechtigten Forscher/-innen auf die Einhaltung von Datenschutzbestimmungen besonders verpflichtet sind (vgl. SUF-Nutzungsrichtlinien des BIBB-FDZ).

<sup>8</sup> Falls auf den Seiten des Metadatenportals des BIBB-FDZ keine Spieldaten zur Verfügung stehen, können selbige auf Anfrage zur Verfügung gestellt werden.

tung von Datenschutzbestimmungen bei Gastaufenthalten verpflichtet werden (vergleiche die Nutzungsrichtlinien für die Gastaufenthalten auf [www.bibb-fdz.de](http://www.bibb-fdz.de)).

Die Mitarbeiter/-innen des BIBB-FDZ nehmen nur zum Zwecke der Beratung, der Verbesserung des BIBB-FDZ-Services sowie zur Gewährleistung der Einhaltung des Datenschutzes Einblick in Forschungsfragen, Methoden und Analysen der Wissenschaftler/-innen. Mitarbeiter/-innen des BIBB, die nicht dem BIBB-FDZ angehören, erhalten keinen Einblick in die Tätigkeiten der Wissenschaftler/-innen.

### 6.3 Anonymisierungen nach Datenprodukten

Die für die einzelnen Datensatzversionen nach Zugangswegen vorgenommenen Anonymisierungen (Vergrößerungen und Löschungen) können der Auflistung im folgenden Abschnitt entnommen werden.

#### 6.3.1 Anonymisierung: SUF

Der Scientific-Use-File (SUF) ist eine jeweils anonymisierte Version des Querschnitt- (Q) und des Ereignisdatensatzes (E) exklusive Textvariablen<sup>9</sup>. Tabelle 7 weist die anonymisierten Variablen im SUF aus.

**Tabelle 7**

Anonymisierte Originalvariablen im SUF der BIBB/BAuA-Erwerbstätigenbefragung 2011/2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008

Variablenname	Variableninhalt	Anonymisierung und Variablennamen (in Klammern)	Datensatz
*ta1005	Ausbildungsberuf der Erstausbildung (KldB 1992 und ISCO-88, ISCO-08 4-Steller sowie KldB 2010 5-Steller)	Vergrößert auf 2-Steller (Kldb92_ta1005_2d Isco88_ta1005_2d Isco08_ta1005_2d Kldb2010_ta1005_2d)	Q
*ta 2101	Erwerbstätigkeit (KldB 1992 und ISCO-88, ISCO-08 4-Steller sowie KldB 2010 5-Steller)	Vergrößert auf 2-Steller (Kldb92_ta2101_2d Isco88_ta2101_2d Isco08_ta2101_2d Kldb2010_ta2101_2d)	E
*ta 2503	Ausbildungsberuf nach Erstausbildung (KldB 1992 und ISCO-88, ISCO-08 4-Steller sowie KldB 2010 5-Steller)	Vergrößert auf 2-Steller (Kldb92_ta2503_2d Isco88_ta2503_2d Isco08_ta2503_2d Kldb2010_ta2503_2d)	E
*ta 2603	berufliche Weiterbildung nach Erstausbildung (KldB 1992 und ISCO-88, ISCO-08 4-Steller sowie KldB 2010 5-Steller)	Vergrößert auf 2-Steller (Kldb92_ta2603_2d Isco88_ta2603_2d Isco08_ta2603_2d Kldb2010_ta2603_2d)	E
*ta 2702	Studium nach Erstausbildung (KldB 1992 und ISCO-88, ISCO-08 4-Steller sowie KldB 2010 5-Steller)	Vergrößert auf 2-Steller (Kldb92_ta2702_2d Isco88_ta2702_2d Isco08_ta2702_2d Kldb2010_ta2702_2d)	E
*ta 3006	Weiterer Ausbildungsberuf (KldB 1992 und ISCO-88, ISCO-08 4-Steller sowie KldB 2010 5-Steller)	Vergrößert auf 2-Steller (Kldb92_ta3006_2d Isco88_ta3006_2d Isco08_ta3006_2d Kldb2010_ta3006_2d)	Q
*ta 3007	Weiterbildungsberuf (KldB 1992 und ISCO-88, ISCO-08 4-Steller sowie KldB 2010 5-Steller)	Vergrößert auf 2-Steller (Kldb92_ta3007_2d Isco88_ta3007_2d Isco08_ta3007_2d Kldb2010_ta3007_2d)	Q

<sup>9</sup> Die Textvariablen stehen für Auswertungen über diesen Datenzugangsweg nicht zur Verfügung. Die Textvariablen können aus Datenschutzgründen nur im Gastaufenthalt im BIBB analysiert werden.

(Fortsetzung Tab. 7)

Variablenname	Variableninhalt	Anonymisierung und Variablennamen (in Klammern)	Datensatz
*ta_3008	Studium (KldB 1992 und ISCO-88, ISCO-08 4-Steller sowie KldB 2010 5-Steller)	Vergrößert auf 2-Steller (Kldb92_ta3008_2d Isco88_ta3008_2d Isco08_ta3008_2d Kldb2010_ta3008_2d)	Q
gkz	Wohnort (Gemeindekennziffer)	gelöscht	-
kkz	Wohnort (Kreiskennziffer)	gelöscht	-
int_tag	Tag des Interviews	gelöscht	-
beruprob_ta*	Variablen zu Problemen bei der Berufevercodung	gelöscht	-

### 6.3.2 Formale Anonymisierung: Datenfernverarbeitung

Spieldaten der schwächer anonymisierten Versionen von Zusatzvariablen (z. B. tiefer gegliederte Regional- oder Berufsangaben) können *im Einzelfall* vom BIBB-FDZ angefordert werden. Voraussetzung dafür ist ein gültiger Nutzungsvertrag für die DFV. Tabelle 8 weist die anonymisierten Variablen der Datenfernverarbeitung aus.

**Tabelle 8**

Vergößerte bzw. nicht verfügbare Variablen für die DFV der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008

Fragebogennr.	Variableninhalt	Anonymisierung	Datensatz
gkz	Wohnort (Gemeindekennziffer)	Kreiskennziffer nach Gebietsstand 2014 (Kkz_11)	Q
kkz	Wohnort (Kreiskennziffer)	Raumordnungsregionen (Ror Ror_systemfrei)	Q
kkz	Wohnort (Kreiskennziffer)	Arbeitsagenturbezirke (aab)	Q
kkz	Wohnort (Kreiskennziffer)	NUTS2 (nuts2)	Q
int_tag	Tag des Interviews	gelöscht	-
beruprob_ta*	Variablen zu Problemen bei der Berufevercodung	gelöscht	-

### 6.3.3 Projektbezogene Anonymisierung: Gastrechnen im BIBB

An den Gastwissenschaftlerarbeitsplätzen im BIBB-FDZ stehen projektbezogene, anonymisierte Versionen des Querschnitt- (Q) und des Ereignisdatensatzes (E) inklusive der Textvariablen zur Verfügung. Tabelle 9 weist die anonymisierten Variablen für einen Gastaufenthalt aus.

Tabelle 9

Anonymisierte Variablen im GWA der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008

Fragebogenr.	Variableninhalt	Anonymisierung	Datensatz
gkz	Wohnort (Gemeindekennziffer)	Kreiskennziffer nach Gebietsstand 2014 (Kkz_11)	Q
kkz	Wohnort (Kreiskenziffer)	Raumordnungsregionen (Ror Ror_systemfrei)	Q
kkz	Wohnort (Kreiskenziffer)	Arbeitsagenturbezirke (aab)	Q
kkz	Wohnort (Kreiskenziffer)	NUTS2 (nuts2)	Q
int_tag	Tag des Interviews	gelöscht	-
beruprob_ta*	Variablen zu Problemen bei der Berufevecodung	gelöscht	-

## 7 Arbeitshilfen

Auf den Internetseiten der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 des BIBB-FDZ und im Metadatenportal (unter <http://metadaten.bibb.de/metadaten/87>) stellen wir verschiedene datensatzspezifische Arbeitshilfen zum Download zur Verfügung. Hierzu gehören neben diesem Datenhandbuch der Fragebogen der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 und der Methodenbericht von TNS Infratest Sozialforschung.

In den folgenden Abschnitten informieren wir über die Bildung der GewichtungsvARIABLEN (7.1) und die Deklaration fehlender Werte nach dem BIBB-FDZ-Standard (7.2).

### 7.1 Gewichtung

Die Gewichtung der BIBB/BAuA-Erwerbstätigenbefragung 2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008 erfolgte zweistufig auf Grundlage des Bildungsberichts „Bildung in Deutschland 2010“ vom Bundesministerium für Bildung und Forschung sowie der Fachserie 11, Reihe 3 „Bildung und Kultur – Berufliche Bildung, 2009“ des Statistischen Bundesamtes. Zunächst wurden die Anteile der Ausbildungsarten an die Referenzstatistik angepasst (Variable: *gew1*) (GENSICKE/TSCHERSICH 2012, S. 14).

Anschließend erfolgte eine Anpassung der Randsummen innerhalb der dualen Ausbildung und der Hochschulausbildung. Dazu wurden folgende Merkmalsausprägungen genutzt: West/Ost und Geschlecht sowie die Wirtschaftsbereiche Handwerk/Industrie und Handel/öffentlicher Dienst für die duale Ausbildung; Geschlecht und Universität/Fachhochschule für die Hochschulausbildung (GENSICKE/TSCHERSICH 2012, S. 14). Dieses Strukturgewicht ist im Datensatz in der Variable *gew2* enthalten. Nähere Informationen zur Gewichtung können dem Feldbericht von TNS Infratest entnommen werden.

### 7.2 Deklaration von Filterwerten und fehlenden Werten nach dem BIBB-FDZ-Standard

In den Forschungsdaten wurden bei allen betreffenden Variablen Filterwerte (–1) neu angelegt, da sie im Ursprungsdatensatz nicht vorhanden waren.

**Tabelle 10**

Standarddeklaration des BIBB-FDZ für verschiedene Arten fehlender Antwortwerte

Wert	Wertelabel
–9	keine Angabe
–8	weiß nicht
–7	trifft nicht zu
–6	ungültig (z. B. Ausschluss durch Datenbereinigung)
–5	verweigert
–1	Filter

In Datensätzen des BIBB-FDZ werden grundsätzlich keine Missing Values definiert. Das entscheiden die Anwender/-innen selbstständig. Eine Ausnahme bildet der Ereignisdatensatz, in welchem Fälle, die nicht zum entsprechenden Spell gehören, Missings aufweisen.

### 7.3 Zitierhinweise

Jede Arbeit, die einen im BIBB-FDZ verfügbaren Datensatz nutzt oder sich darauf bezieht, sollte eine Berücksichtigung dieser Quelle in Form einer bibliografischen Zitation im Literaturverzeichnis oder in der Fußnote enthalten. Zusätzlich zu der Zitierung der Daten selbst sollte gegebenenfalls auf die entsprechenden BIBB-FDZ Daten- und Methodenberichte verwiesen werden.

Ausführliche Informationen hierzu finden Sie auf unserem Metadatenportal.

# Literatur

GENSICKE, Miriam; TSCHERSICH, Nikolai (2012): BIBB/BAuA-Erwerbstätigenbefragung 2011/2012 – Ausbildungsabsolventen der Jahre 2006 bis 2008. Feldbericht – Los 3. TNS Infratest Sozialforschung, München.







Bundesinstitut für Berufsbildung  
Robert-Schuman-Platz 3  
53175 Bonn

Telefon: (0228) 107-0  
Telefax: (0228) 107 2976/77

Internet: [www.bibb.de](http://www.bibb.de)  
E-Mail: [zentrale@bibb.de](mailto:zentrale@bibb.de)

Bundesinstitut  
für Berufsbildung **BiBB** ▶

- ▶ Forschen
- ▶ Beraten
- ▶ Zukunft gestalten