

Forschungsdatenzentrum

der Bundesagentur für Arbeit  
im Institut für Arbeitsmarkt-  
und Berufsforschung

FDZ

# FDZ-Datenreport

Dokumentation zu Arbeitsmarktdaten

11/2016

DE

## IAB-SOEP Migrationsstichprobe verknüpft mit administrativen Daten des IAB

### Version 1

(IAB-SOEP-MIG-ADIAB 7514, Version 1)

Parvati Trübswetter,  
Tanja Fendel



Bundesagentur für Arbeit

# IAB-SOEP Migrationsstichprobe verknüpft mit administrativen Daten des IAB, Version 1 (IAB-SOEP-MIG-ADIAB 7514, Version 1)

Parvati Trübswetter, Tanja Fendel  
(Institut für Arbeitsmarkt- und Berufsforschung)

Datensatz: IAB-SOEP-MIG-ADIAB 7514, Version 1  
Dokumentation: IAB-SOEP-MIG-ADIAB 7514 DE v1 dok1

Die FDZ-Datenreporte beschreiben die Daten des FDZ im Detail. Diese Reihe hat somit eine doppelte Funktion: zum einen stellen Nutzerinnen und Nutzer fest, ob die angebotenen Daten für das Forschungsvorhaben geeignet sind, zum anderen dienen sie zur Vorbereitung der Auswertungen.

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of reports has a dual function: on the one hand, those using the reports can ascertain whether the data offered is suitable for their research task; on the other, the data can be used to prepare evaluations.

## Inhaltsverzeichnis

<b>1</b>	<b>EINLEITUNG UND KURZBESCHREIBUNG .....</b>	<b>9</b>
1.1	Einleitung.....	9
1.2	<b>Datennutzung.....</b>	<b>10</b>
	Datenzugang und Datenhaltung .....	10
	Dateinamen der administrativen Daten.....	13
	Dateinamen der IAB-SOEP Migrationsstichprobe.....	13
1.3	<b>Steckbrief.....</b>	<b>14</b>
1.4	<b>Merkmalsliste der administrativen Daten.....</b>	<b>18</b>
1.5	<b>Mengengerüst .....</b>	<b>22</b>
<b>2</b>	<b>DATENQUELLEN .....</b>	<b>22</b>
2.1	IAB-SOEP Migrationsstichprobe .....	22
2.2	Administrative Daten aus den IEB.....	23
2.3	Beschäftigtenhistorik (BeH) .....	23
2.4	Leistungsempfängerhistorik (LeH) .....	24
2.5	Leistungshistorik Grundsicherung (LHG).....	25
2.6	Arbeitsuchendenhistorik (ASU/XASU) .....	26
2.7	Maßnahmeteilnahmehistoriken (MTH).....	26
<b>3</b>	<b>DATENAUFBEREITUNG UND STICHPROBENZIEHUNG .....</b>	<b>26</b>
3.1	<b>Bereinigungen und Aufbereitungen.....</b>	<b>26</b>
	Beschäftigtenhistorik (BeH) .....	27
	Leistungsempfängerhistorik (LeH).....	27
	Leistungshistorik Grundsicherung (LHG).....	28
	Arbeitsuchendenhistorik ASU und XASU .....	28
	Maßnahmeteilnahmehistoriken (MTH).....	29
	SGB-II-Anonymisierung.....	30
3.2	<b>Episodensplitting .....</b>	<b>30</b>
3.3	<b>Stichprobenziehung.....</b>	<b>31</b>
3.4	<b>Fehlende Werte .....</b>	<b>32</b>
<b>4</b>	<b>DATENVERKNÜPFUNG .....</b>	<b>32</b>
4.1	<b>Bereinigungen und Standardisierung personenbezogener Merkmale .....</b>	<b>33</b>

4.2	Linkage der Ankerpersonen .....	34
4.3	Exaktes Record Linkage .....	34
4.4	Probabilistisches Record Linkage .....	35
<b>5</b>	<b>DATENQUALITÄT UND PROBLEME .....</b>	<b>35</b>
5.1	IAB-SOEP Migrationsstichprobe .....	35
5.2	Gesamte IEB .....	35
5.3	Beschäftigtenhistorik (BeH) .....	37
5.4	Leistungsempfängerhistorik (LeH) .....	39
5.5	Leistungshistorik Grundsicherung (LHG).....	39
5.6	Arbeitsuchendehistorik (ASU/XASU) .....	40
	ASU .....	40
	XASU.....	41
5.7	Maßnahmeteilnahmehistoriken (MTH).....	41
<b>6</b>	<b>VARIABLENBESCHREIBUNG .....</b>	<b>42</b>
6.1	Identifikatoren .....	42
	Systemfreie Personennummer (persnr).....	42
	Systemfreie Betriebsnummer (betnr) .....	43
6.2	Generierte technische Merkmale .....	44
	Satzzähler pro Konto (spell) .....	44
	Quelle des Satzes (quelle).....	44
	Jahr (jahr) .....	44
6.3	Gültigkeitszeitraum .....	45
	Beginndatum Originalbeobachtung (begorig) .....	45
	Enddatum Originalbeobachtung (endorig).....	45
	Beginndatum der Episode (begepi) .....	46
	Enddatum der Episode (endept) .....	46
6.4	Informationen zur Person .....	46
	Geschlecht (frau).....	46
	Geburtsjahr (gebjahr).....	47
	Staatsangehörigkeit (nation).....	47
	Staatsangehörigkeit vergrößert (nation_gr).....	47
	Familienstand (famst) .....	48
	Kinderzahl (kind).....	48
	Ausbildung (ausbildung).....	49
	Schulausbildung (schule) .....	50
6.5	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche .....	51
	Tagesentgelt/ täglicher Leistungssatz (tentgelt) .....	51
	Beruf – ausgeübte/letzte Tätigkeit (KIdB 1988) (beruf) .....	52
	Berufsgruppe – ausgeübte/letzte Tätigkeit (KIdB 2010) (beruf2010_3).....	53

Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_4) .....	54
Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau) .....	55
Teilzeit (teilzeit) .....	55
Erwerbsstatus (erwstat) .....	56
Gleitzone (gleitz) .....	58
Leiharbeit (leih) .....	58
Befristung (befrist) .....	59
Abmeldegrund / Abgabegrund / Beendigungsgrund (grund) .....	59
Erwerbsstatus vor Arbeitsuche (estatvor) .....	60
Status nach Arbeitsuche (estatnach) .....	61
Profillage (profil) .....	61
Art der Kündigung der letzten Tätigkeit (art_kuend) .....	62
Arbeitszeit des Stellengesuchs (arbzeit) .....	62
Restanspruch/geplante Dauer (restanspruch) .....	62
Trägerart (traeger) .....	63
Beginndatum der Arbeitslosigkeit (alo_beg) .....	63
Dauer der Arbeitslosigkeit (alo_dau) .....	64
<b>6.6 Ortsangaben .....</b>	<b>64</b>
Wohnort - Kreis (wo_kreis) .....	64
Wohnort - Bundesland (wo_bula) .....	65
Wohnort - Arbeitsagentur (wo_aa) .....	65
Wohnort - Regionaldirektion (wo_rd) .....	66
<b>6.7 Betriebsmerkmale .....</b>	<b>66</b>
WZ73 3-Steller (w73_3) .....	66
WZ93 5-Steller (w93_5) .....	67
WZ93 3-Steller (w93_3) .....	67
WZ03 5-Steller (w03_5) .....	68
WZ03 3-Steller (w03_3) .....	68
WZ08 5-Steller (w08_5) .....	69
WZ08 3-Steller (w08_3) .....	69
W73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen) .....	70
Art der Vervollständigung w73_3 (group_w73_3) .....	70
W93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen) .....	70
Art der Vervollständigung w93_3 (group_w93_3) .....	71
Jahr des 1. Auftretens der Betriebsnummer (grd_jahr) .....	71
Erstes Auftreten Betriebsnummer (grd_dat) .....	72
Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr) .....	72
Letztes Auftreten Betriebsnummer (lzt_dat) .....	73
Anzahl Beschäftigte gesamt (az_ges) .....	73
Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz) .....	73
Anzahl geringfügig Beschäftigte (az_gf) .....	74
Mittelwert imputiertes Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw) .....	74
Arbeitsort Kreis (ao_kreis) .....	75
Arbeitsort Bundesland (ao_bula) .....	75
<b>6.8 Verknüpfungsmerkmale .....</b>	<b>75</b>
Zustimmung zur Verknüpfung (match_ev) .....	75
Vorliegen administrativer Personendaten (match_admin) .....	76
Anzahl der Treffer in den administrativen Daten pro befragter Person (match_az) .....	76
Verknüpfungsmethode (match_typ) .....	76
Qualität der Verknüpfung (match_qual) .....	77
<b>7 LITERATUR .....</b>	<b>78</b>

<b>8</b>	<b>ANHANG.....</b>	<b>80</b>
8.1	Auszählungen .....	80
<b>9</b>	<b>ABKÜRZUNGSVERZEICHNIS.....</b>	<b>81</b>

## **Zusammenfassung**

Dieser Datenreport beschreibt die verknüpften Befragungsdaten der IAB-SOEP Migrationsstichprobe mit administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB).

## **Abstract**

This data report describes the linked survey data of the IAB-SOEP Migration Sample with administrative data of the Institute for Employment Research (IAB).

**Keywords:** German administrative micro data, labour market data, data manual, Migration data, survey data

Für ihre Mitarbeit und Unterstützung danken wir den Kollegen des Bereichs Internationale Vergleiche und Europäische Integration und des Forschungsdatenzentrums (FDZ) des Instituts für Arbeitsmarkt- und Berufsforschung. Der vorliegende Datenreport enthält die Beschreibung der IAB-SOEP Migrationsstichprobe verknüpft mit administrativen Daten des IAB, Version 1.

## **Datenverfügbarkeit**

Der in diesem Beitrag beschriebene Datensatz ist für die Fachöffentlichkeit zugänglich. Nähere Informationen dazu auf der Internetseite <http://fdz.iab.de/>.



# 1 Einleitung und Kurzbeschreibung

## 1.1 Einleitung

Dieser Datenreport beschreibt die verknüpften Befragungsdaten der IAB-SOEP Migrationsstichprobe mit administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB). Diese verknüpften Prozess- und Befragungsdaten bestehen aus in Interviews erhobenen Informationen von ausgewählten Migranten, die in der ersten und/ oder der zweiten Welle der IAB-SOEP Migrationsstichprobe enthalten sind und ihren administrativen Personendaten aus den Integrierten Erwerbsbiographien (IEB), sofern sie ihre Zustimmung zur Verknüpfung der Daten erteilt haben und zugeordnet werden konnten.

Zielgruppe der Befragung sind Personen mit Migrationshintergrund im Sinne der Definition des Statistischen Bundesamtes, also Personen, die selbst zugewandert sind, die eine ausländische Staatsangehörigkeit haben oder bei denen mindestens ein Elternteil zugewandert ist oder eine ausländische Staatsangehörigkeit besitzt. Bei der Durchführung des Projektes kooperiert das IAB mit dem Sozio-ökonomischen Panel (SOEP) des Deutschen Instituts für Wirtschaftsforschung (DIW). Beide Forschungseinrichtungen bringen in gleichem Umfang Mittel für die Erhebung der Stichprobe ein. Der Teil des IAB wurde zur Hälfte vom Bundesministerium für Arbeit und Soziales (BMAS) finanziert.

Die befragten Personen wurden aus dem Kreis aller Individuen ausgewählt, die erstmals nach 1994 in den IEB auftauchten. Für die Stichprobenziehung erfolgte zunächst eine Gruppierung verfügbarer Adressen von ca. 2.500 Personen in geografische Stichproben-Regionen. Als Zielpersonen galten zum einen diejenigen, die in den Daten eindeutig als ausländisch identifiziert werden, bspw. aufgrund einer nicht-Deutschen Staatsangehörigkeit und auch diejenigen, die an spezifischen Maßnahme für Migranten teilgenommen haben. Zum anderen wurden für die Identifikation von Personen mit Migrationshintergrund onomastische Verfahren (sprachliche Analysen von Personennamen) angewendet. Die Stichprobe für die Befragung wurde schließlich auf Grundlage soziodemografischer Merkmale aus den IEB gezogen.

Die administrativen Daten stammen aus den Integrierten Erwerbsbiografien (IEB) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB). Sie enthalten im vorliegenden Fall Daten zu allen Personen, die im Rahmen der IAB-SOEP Migrationsstichprobe befragt wurden, einer Verknüpfung zugestimmt haben und die im Beobachtungszeitraum mindestens einmal einen der folgenden Zustände aufweisen:

- sozialversicherungspflichtige Beschäftigung in Deutschland (erfasst ab 1975),
- geringfügige Beschäftigung in Deutschland (erfasst ab 1999),

- Bezug von Leistungen nach dem Rechtskreis SGB III (erfasst ab 1975) oder SGB II (erfasst ab 2005),
- bei der Bundesagentur für Arbeit (BA) bzw. den Grundsicherungsträgern als arbeitsuchend gemeldet (erfasst ab 2000),
- Teilnahme an einer arbeitsmarktpolitischen Maßnahme der BA (erfasst ab 2000).

Diese aus unterschiedlichen Datenquellen stammenden Informationen werden in den IEB zusammengeführt, tagesgenau abgebildet und für die Nutzer aufbereitet.

Während die Beschäftigtenhistorik (BeH) Ursprung der Informationen zu sozialversicherungspflichtiger und geringfügiger Beschäftigung ist, wird der Bezug von Leistungen nach den Rechtskreisen des SGB III und des SGB II in der Leistungsempfängerhistorik (LeH) bzw. in der Leistungshistorik Grundsicherung (LHG) erfasst. Die Arbeitsuchendenhistoriken (ASU und XASU) stellen die Datenquelle für die durch die BA bzw. die durch die kommunalen Grundsicherungsträger erfassten Zeiten der Arbeitsuche dar, während die Teilnahme an arbeitsmarktpolitischen Maßnahmen in der Maßnahmeteilnahmehistorik (MTH) erfasst wird. Anders als in früheren Versionen der SIAB sind Informationen zur Teilnahme an arbeitsmarktpolitischen Maßnahmen aus der MTH in der aktuellen Version der SIAB ebenfalls enthalten.

Die IAB-SOEP-MIG-ADIAB wird im Forschungsdatenzentrum der Bundesagentur für Arbeit im IAB produziert.

Dieser Datenreport gliedert sich wie folgt. Neben der Einleitung enthält Kapitel 1 unter anderem Informationen zum Datenzugang sowie eine Kurzbeschreibung der Daten, das Mengengerüst und eine Merkmalsliste. Eine Beschreibung der einzelnen Datenquellen folgt in Kapitel 2. Die Datenaufbereitung sowie die Stichprobenziehung werden in Kapitel 3 diskutiert, Kapitel 4 und 5 enthalten eine Beschreibung der Datenverknüpfung sowie der Datenqualität. In Kapitel 6 werden die einzelnen Merkmale dargestellt.

## **1.2 Datennutzung**

### **Datenzugang und Datenhaltung**

Die IAB-SOEP MIG ADIAB-Daten sind schwach anonymisiert und können ausschließlich im Rahmen von Gastaufenthalten am FDZ und anschließender Datenfernverarbeitung ausgewertet werden.

Für die Datennutzung ist zunächst ein Antrag beim FDZ einzureichen, über die Genehmigung des Forschungsprojekts entscheidet das Bundesministerium für Arbeit und Soziales (BMAS). Nach erfolgter Genehmigung wird ein Nutzungsvertrag mit der Institution des Forschers geschlossen. Details zur Beantragung des Datensatzes und Datenverarbeitungsmöglichkeiten können der FDZ-Website entnommen werden.

Die IAB-SOEP Migrationsstichprobe ist – entsprechend dem SOEP - nach Themen gegliedert und in mehreren Dateien abgelegt. Sie enthalten zunächst die im persönlichen Interview erhobenen Merkmale der Befragten im Original sowie deren Personen- und Haushaltsidentifikatoren. Es gibt weiterhin Dateien, die bereinigte und neu gebildete Variablen enthalten und Dateien mit biographischen Inhalten. Technische und Hintergrundinformationen werden in weiteren Dateien angeboten. Alle Datensätze können über persönliche und Haushaltsindikatoren zusammengespielt werden (siehe Dokumentation auf <https://paneldata.org/studies/6>, eine Variablenbeschreibung findet sich für den Hauptdatensatz der Welle 1 unter [https://data.soep.de/soep-core/data/bdp\\_mig](https://data.soep.de/soep-core/data/bdp_mig) und für Welle 2 unter [https://data.soep.de/soep-core/data/bep\\_mig](https://data.soep.de/soep-core/data/bep_mig)). Die Variablen der weiteren Datensätze können aufgelistet werden, indem `bdp_mig` im Link oben durch den entsprechenden Dateinamen ersetzt wird.

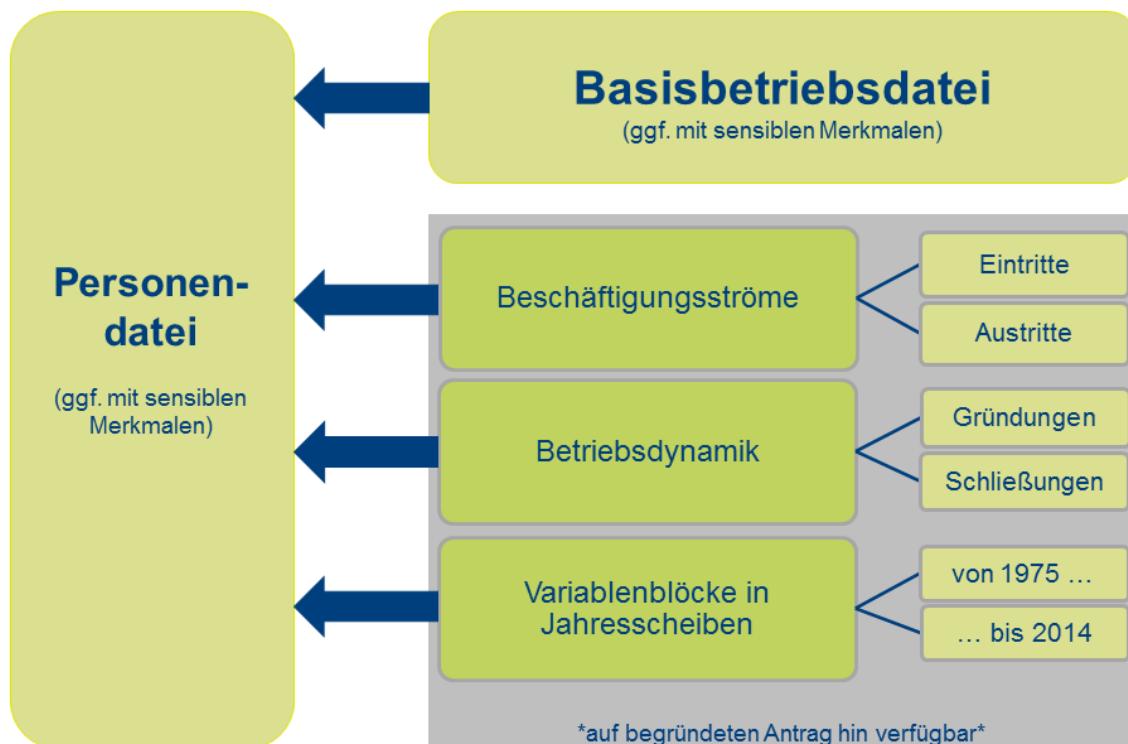
Ein weiteres Element der IAB-SOEP-MIG-ADIAB ist die Verknüpfungsdatei, welche die Personnummer sowie technische Merkmale enthält, die Informationen darüber enthalten, ob bzw. wie die Verknüpfung mit den administrativen Daten gelungen ist. Die Datei enthält auch die Befragten aus der IAB-SOEP Migrationsstichprobe, zu denen keine verknüpften administrativen Daten gefunden werden konnten. Somit können eigene Ausfallanalysen auf Basis aller Teilnehmer der IAB-SOEP Migrationsstichprobe durchgeführt werden.

Die administrativen Daten, die sowohl deutsche als auch englische Labels enthalten,<sup>1</sup> sind modular gegliedert und in mehreren Dateien abgelegt. Ein Modul, das im weiteren Verlauf Personendatei genannt wird, enthält Identifikatoren (systemfreie Personen- und Betriebsnummer), persönliche Merkmale, Informationen zu Beschäftigung, Leistungsbezug, Arbeitsuche und Teilnahme an arbeitsmarktpolitischen Maßnahmen, Merkmale zum Wohnort sowie technische Merkmale. Ein zweites Modul, die Basisbetriebsdatei, enthält die Betriebsnummer, das Jahr, Merkmale zu Arbeitsort und Wirtschaftszweigen sowie weitere Betriebsinformationen zum Stichtag 30. Juni. Für Ostdeutschland liegen die Betriebsdaten erst ab 1992 vor. Somit sind Betriebs- und Personenmerkmale organisatorisch voneinander getrennt, was die Struktur der Daten deutlich macht und Speicherplatz spart (vgl. Abbildung 1). Die Verknüpfung von Basisbetriebs- und Personendatei erfolgt über die Betriebsnummer und das Jahr des Datensatzes. Die in der Merkmalsliste (vgl. S. 18ff) mit einem „\*“ gekennzeichneten Variablen, sind in der Basisbetriebsdatei enthalten. Zu beachten ist, dass das für die Verknüpfung beider Module notwendige Merkmal „Betriebsnummer“ sowohl in der Personen- als auch in der Betriebsdatei vorhanden ist.

---

<sup>1</sup> Mit den Stata-Befehlen *label language en* oder *label language de* kann auf englisch- bzw. deutschsprachige Labels umgestellt werden.

Abbildung 1 Datenhaltung der SIAB-Daten



Bestimmte Merkmale, die eine Deanonymisierung von Personen oder Betrieben erleichtern, werden nur dann im Original weitergegeben, wenn es für das Analyseziel notwendig ist und dies im Antrag auf den Datenzugang explizit begründet wird. Diese aus datenschutzrechtlicher Sicht besonders sensiblen Merkmale sind:

Personendatei:

- Staatsangehörigkeit (nation)
- Berufsuntergruppe (beruf2010\_4)
- Wohnort Arbeitsagentur (wo\_aa)
- Wohnort Kreis (wo\_kreis)

Basisbetriebsdatei:

- Arbeitsort Kreis (ao\_kreis)
- Wirtschaftszweig - Wirtschaftsunterklasse (w93\_5)
- Wirtschaftszweig - Wirtschaftsunterklasse (w03\_5)
- Wirtschaftszweig - Wirtschaftsunterklasse (w08\_5)
- Erstes Auftreten Betriebsnummer (grd\_dat)
- Letztes Auftreten Betriebsnummer (lzt\_dat)

Auf begründeten Antrag können aus dem Betriebs-Historik-Panel (BHP) weitere Betriebsmerkmale in Variablenblöcken bereitgestellt werden (vgl. Abbildung 1). Nähere Informationen zum BHP finden sich unter <http://fdz.iab.de/>.

### **Dateinamen der administrativen Daten**

#### Personendatei

IAB-SOEP-MIG-ADIAB\_7514\_v1.dta

#### Basisbetriebsdatei

IAB-SOEP-MIG-ADIAB\_7514\_v1\_bhp\_basis\_v1.dta

#### Verknüpfungsdatei

IAB-SOEP-MIG-ADIAB\_7514\_v1\_key.dta

#### Zusatzmodule BHP

Bestandsdaten mit Variablenblöcken

IAB-SOEP-MIG-ADIAB\_7514\_v1\_bhp\_v1\_####.dta, #### = 1975 - 2014

Beschäftigtenströme

IAB-SOEP-MIG-ADIAB\_7514\_v1\_bhp\_inflow\_v1.dta

IAB-SOEP-MIG-ADIAB\_7514\_v1\_bhp\_outflow\_v1.dta

Betriebsdynamik

IAB-SOEP-MIG-ADIAB\_7514\_v1\_bhp\_entry\_v1.dta

IAB-SOEP-MIG-ADIAB\_7514\_v1\_bhp\_exit\_v1.dta

### **Dateinamen der IAB-SOEP Migrationsstichprobe**

<b>Datensatz</b>	<b>Beschreibung / Dokumentation des Files</b>
bdh, beh	Alle Variablen aus dem Haushaltsfragebogen
bdhbrutto, behbrutto	Grundinformationen zu allen Haushalten, die kontaktiert werden sollten
bdhgen, behgen	Generierte Variablen aus der Haushaltsbefragung
bdkind, bekind	Variablen zu Kindern
bdp, bep	Die kompatiblen Variablen aus dem integrierten Personen- und Biografiefragebogen, jedoch ohne die Biografie-Anteile
bdp_mig, bep_mig	Alle Variablen aus dem integrierten Personen- und Biografiefragebogen der IAB-SOEP Migrationsstichprobe (Erstbefragte).
bdpbrutto, bepbrutto	Grundinformationen zu allen Personen, die in einem SOEP-Haushalt leben
bdpgen, bepgen	generierte Variablen aus dem Personenfragebogen
bdpkal, bepkal	Kalendarium zu Erwerbstätigkeit und Einkommen
biobirth	Geburtenbiografie der weiblichen Befragten

biocouplm	Paargeschichte: Monatsspells
biocouply	Paargeschichte: Jahresspells
bioedu	Daten zu Bildungsbeteiligung und Bildungsübergängen
bioimmig	Längsschnittvariablen zu Einwanderern und Ausländern (jährliche Daten)
biomarsm	Heiratsgeschichte: Monatsspells
biomarsy	Heiratsgeschichte: Jahresspells
bioparen	Biografische Daten zu Eltern
biosib	Geschwistergeschichte
biosoc	Retrospektive Daten zu Jugend und Sozialisation
biotwin	Daten zu Zwillingen
cirdef	Datensatz zur Auswahl von Zufallsgruppen je Primary Sampling Unit, z.Bsp. zur Erstellung einer Lehrstichprobe (50%)
design	Designgewichte (haushaltsbezogen)
hpfad	Surveybezogene Variablen zu allen Haushalten, die interviewt wurden
kidlong	Long-Datensatz zu den wellenspezifischen \$\$kind-Datensätzen
lifespell	Informationen zur pre- und post-survey-Geschichte für alle Personen, die jemals Mitglied eines SOEP-Haushalts waren
migspell	Migrationsbiografien aus bdp_mig als Spelldatensatz
mihinc	Ergebnisse der multiplen Imputation des Nettoeinkommens des Haushalts
pbiospe	Jährliche Aktivitätsbiografie (retrospektiv, dann prospektiv fortgeschrieben)
pflege	Variablen zu Schwerbehinderung und Pflegebedürftigkeit
ppfad	Basis-Informationen und surveybezogene Variablen zu allen Personen, die in Teilnehmerhaushalten, inkl. der Personen, die nicht erreichbar waren oder verweigerten.
artkalen	Beschäftigungskalender auf Monatsbasis
bepage17	Fragen aus dem Jugendfragebogen, die nicht in bioage17.dta enthalten sind
beschool	Fragebogen über 12-13 Jährige im Haushalt
bevpt	Fragebogen über verstorbene Person im Haushalt
bioage17	Generierte biographische Information aus dem Jugendfragebogen
bioagel	Generierte biographische Information
biojob	Informationen über den ersten und letzten Job
cogdj	Daten über kognitive Test (Jugend)
hbrutt13	Originale Bruttostichprobe der ersten Welle der IAB-SOEP MIG
health	Variablen zum Gesundheitszustand
hhf	Gewichtungsvariablen Haushaltsebene
pbr_exit	Kumulierte Exit-Variablen
phrf	Gewichtungsvariablen Personenebene

### 1.3 Steckbrief

Tabelle 1 Steckbrief IAB-SOEP-MIG-ADIAB

Themen/ Merkmalsgruppen	<u>Befragungsdaten:</u> <b>Migrationsbiografie:</b> Einreisezeitpunkt; frühere Migrationsepisoden; Such- und Informationsverhalten; soziale Netzwerke; <b>Bildungsbiografie:</b> höchste Bildungs- und Berufsabschlüsse; Schuljahre; Bildungserwerb im In- und Ausland; Anerkennungsverfahren; Sprachkompetenz; <b>Erwerbsbiografie:</b> Beschäftigung; Selbständigkeit; Arbeitslosigkeit in Deutschland und dem Herkunftsland;
-------------------------	--

	<p><b>Arbeitsmarktinformationen:</b> Verdienste; Voll- und Teilzeitarbeit; Arbeitsstunden; Leistungsbezug; Reservationslohn; Teilnahme an Maßnahmen der aktiven Arbeitsmarktpolitik;</p> <p><b>Rückkehrmigration:</b> Rückkehrmigrationsabsichten; Rückkehrmigration; eingeschränkte Befragung von Rückkehrmigranten;</p> <p><u>Administrative Personendaten</u></p> <p><b>Beschäftigtenhistorik (BeH):</b> Jahres- und Abmeldungen zur Sozialversicherung für sozialversicherungspflichtig und geringfügig Beschäftigte</p> <p><b>Leistungsempfängerhistorik (LeH):</b> Informationen zu Leistungsbezug nach SGB III für Bezieher von Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld</p> <p><b>Leistungshistorik Grundsicherung aus A2LL (LHG):</b> Daten zu Personen, die Leistungen zur Sicherung des Lebensunterhalts nach SGB II beziehen (Trägerarten: ARGE/gemeinsame Einrichtungen, getrennte Trägerschaft/getrennte Aufgabenwahrnehmung, zugelassene kommunale Träger)</p> <p><b>Arbeitsuchendehistorik (ASU):</b> Informationen zur Arbeitsuche</p> <p><b>Arbeitsuchendehistorik aus XSozial-BA-SGB II (XASU):</b> Informationen zur Arbeitsuche, die über den Übermittlungsstandard XSozial-BA-SGB II von zugelassenen kommunalen Trägern an die BA gemeldet werden.</p> <p><b>Maßnahmeteilnehnehistoriken (MTH):</b> Informationen zu Teilnahmen an Maßnahme der aktiven Arbeitsmarktförderung (keine Maßnahmen der zugelassenen kommunalen Träger im SGB II)</p> <p><u>Administrative Betriebsdaten</u></p> <p><b>Kerndatensatz:</b> Betriebsmerkmale (z.B. systemfreie Betriebsnummer, Wirtschaftszweig, Bundesland), Erwerbsstruktur der Beschäftigten (z.B. Anzahl der Beschäftigten nach Geschlecht, Art der Beschäftigung), Altersstruktur der Beschäftigten, Struktur der Beschäftigten nach Schul- und Berufsausbildung, Struktur der Beschäftigten nach Berufsklassifikation (Blossfeld Klassifikation), Durchschnittslöhne der Vollzeitbeschäftigten</p> <p><b>Erweiterungsmodul- Beschäftigtenströme:</b> Anzahl Zugänge gesamt, Struktur der Zugänge nach Geschlecht, Beschäftigungsart, Berufsklassifikation (Blossfeld Klassifikation), Wiedereinstellung, Alter, Anzahl Abgänge gesamt, Struktur der Abgänge nach Geschlecht, Beschäftigungsart, Berufsklassifikation (Blossfeld Klassifikation), Betriebszugehörigkeit, Alter</p> <p><b>Erweiterungsmodul- Betriebsdynamik:</b> Gründungsjahr, Art der Gründung, Hilfsvariablen zur Klassifizierung der Gründung, Schließungsjahr, Art der Schließung, Hilfsvariablen zur Klassifizierung der Schließung</p>
Untersuchungseinheit	<p><u>Befragungsdaten:</u> Ankerpersonen mit Migrationshintergrund aus den IEB und deren Haushaltsmitglieder</p> <p><u>Administrative Personendaten:</u> Sozialversicherungspflichtig Beschäftigte (ab 1999 auch geringfügig Beschäftigte), Leistungsempfänger, Arbeitssuchende, Maßnahmeteilnehmer</p>
Fallzahlen	<p><u>Befragungsdaten:</u> Welle 1: 4.964 Personen in 2.723 Haushalten, davon sind 1.543 Personen verknüpft.</p>



	<p>Welle 2: 3752 Personen in 2021 Haushalten, davon sind zusätzlich 510 Personen verknüpft.</p> <p><u>Administrative Daten:</u>  2053 Personen mit 2066 Konten  60.789 Originalsätze  108.919 überschneidungsfreie Sätze (nach Episodensplitting)</p>
Zeitraum	<p><u>Befragungsdaten:</u>  Welle 1: 2013  Welle 2: 2014</p> <p><u>Administrative Daten:</u>  Der Zeitraum hängt von der Datenquelle ab.  BeH: 01.01.1975 - 31.12.2014 (2012: 30-Monats-Datei, 2013: 18-Monats-Datei; 2014: 6-Monats-Datei)  LeH: 29.11.1976 - 31.12.2014  ASU: 07.11.1994 - 31.12.2014  LHG: 01.01.2005 - 31.12.2014  XASU: 01.01.2005 - 31.12.2014  MTH: 02.11.1993 - 31.12.2014</p>
Zeitbezug	<p><u>Befragungsdaten:</u>  Befragungszeitpunkt</p> <p><u>Administrative Personendaten:</u>  Tagesgenaue Erwerbsbiografien</p> <p><u>Administrative Betriebsdaten:</u>  Stichtag 30.Juni</p>
Regionale Gliederung	Bundesländer, Kreise
Gebietsstand	Gebietsstandsbereinigt zum 31.12.2014
Erhebungsdesign	<p><u>Befragungsdaten:</u>  Welle 1: Die Stichprobe ist aus allen Personen in den IEB Daten gezogen worden, die nach 1994 dort das erste Mal erscheinen. Es wurden Cluster mit je 2.500 Personen mit Hilfe von geokodierten Daten des FDZ gebildet, welche dann insgesamt 6.725 geografische Stichprobeneinheiten (PSUs) ergeben. Im nächsten Schritt wurden alle Personen mit Migrationshintergrund anhand der Nationalität und des Onomastik Verfahrens identifiziert. Das Onomastik Verfahren wurde nur auf eine Teilstichprobe von 600 aus insgesamt 6.725 PSUs durchgeführt. Die Auswahl der PSUs erfolgte auf Grundlage geschichteter Zufallsstichproben (stratified random sampling). Nach dem Onomastik-Verfahren wurden alle Personen ohne Migrationshintergrund aus den Daten entfernt. Insgesamt wurden 250 aus den 600 PSUs für die Felderhebung berücksichtigt. In den ausgewählten 250 PSUs besteht die Bruttostichprobe aus 80 Adressen (Haushalte). Bei der Stichprobenziehung wurden entfernungsabhängige und simulierte random walk Verfahren implementiert. Durch ein sogenanntes Screening-Interview mit den Ankerpersonen wurden die Haushalte identifiziert, die die Kriterien des Stichprobendesigns erfüllen.  Welle 2: Haushalte aus der ersten Welle und ausgegründete Haushalte werden befragt, neue Haushaltsmitglieder werden auch aufgenommen.</p> <p><u>Administrative Daten:</u>  Identifikation der Befragten in den Integrierten Erwerbsbiografien des IAB</p>
Beteiligte Institutionen	<p><u>Befragungsdaten:</u>  Auftraggeber: Institut für Arbeitsmarkt- und Berufsforschung in Kooperation mit dem SOEP im DIW Berlin  Durchführung der Befragung: TNS Infratest Sozialforschung</p>



	<u>Administrative Daten:</u> Sozialversicherungsträger, Bundesagentur für Arbeit
Frequenz der Datensammlung	<u>Befragungsdaten</u> Jährlich <u>Administrative Daten</u> Laufend
Dateiformat/ -größe	<u>Befragungsdaten</u> Stata; Personen- und Haushaltsdatensätze für Welle 1, Welle 2 und wellenübergreifende Informationen, Größen der einzelnen Dateien bis zu 9 MB <u>Administrative Daten</u> Stata; Personendatensatz: 14,1 MB; Betriebsdatensatz: 1,88 MB
Dateiorganisation	<u>Befragungsdaten</u> Die Daten werden in mehrere Dateien abgelegt, die wellenspezifische bzw. wellenübergreifende Personen- oder Haushaltsdaten enthalten. <u>Administrative Daten</u> Die Daten werden in zwei Dateien abgelegt. Die eine enthält personenbezogene und die andere betriebsbezogene Informationen. Auf begründeten Antrag werden weitere Dateien mit zusätzlichen Betriebsinformationen bereitgestellt.
Datenzugang	Gastaufenthalt, Datenfernverarbeitung nach vorangegangenen Gastaufenthalt
Anonymisierungsgrad	Schwach anonymisiert
Sensible Merkmale	Staatsangehörigkeit (nation), Berufsuntergruppe (beruf2010_4), Wohnort – Kreis (wo_kreis), Wohnort – Arbeitsagentur (wo_aa), Arbeitsort - Kreis (ao_kreis), Wirtschaftsunterklasse (w93_5), Wirtschaftsunterklasse (w03_5), Wirtschaftsunterklasse (w08_5), Erstes Auftreten der Betriebsnummer (grd_dat), Letztes Auftreten der Betriebsnummer (lzt_dat)
Zitierung der Daten und Datendokumentation	IAB-SOEP MIG ADIAB Welle 1 und 2, Nürnberg 2016 <u>Daten:</u> „Die Datengrundlage dieses Beitrags bildet die IAB-SOEP Migrationsstichprobe verknüpft mit administrativen Daten des IAB (IAB-SOEP MIG ADIAB, Version 1). Der Datenzugang erfolgte über einen Gastaufenthalt am Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ) und/oder mittels kontrollierter Datenfernverarbeitung beim FDZ.“ <u>Datendokumentation:</u> Trübswetter Parvati, Tanja Fendel (2016): IAB-SOEP MIG ADIAB 1975 - 2014. FDZ-Datenreport, 11/2016 (de), Nürnberg.  Herbert Brücker, Martin Kroh, Simone Bartsch, Klaudia Erhardt, Jan Goebel, Simon Kühne, Elisabeth Liebau, Parvati Trübswetter, Ingrid Tucci & Jürgen Schupp (2014): The new IAB-SOEP Migration Sample: an introduction into the methodology and the contents. <a href="#">SOEP Survey Paper 216</a>
Datensatzversion	IAB-SOEP Migrationsstichprobe verknüpft mit administrativen Daten des IAB (IAB-SOEP-MIG-ADIAB 7514, Version 1)

## 1.4 Merkmalsliste der administrativen Daten

Die Merkmalsübersicht in Tabelle 4 listet die Variablennamen und die längeren Merkmalsbezeichnungen auf. Weiterhin bietet sie eine Übersicht darüber, welche Merkmale für welche Datenquellen gefüllt sind:

**Tabelle 2 Füllgrad der Merkmale**

	Merkmal ist für die Quelle gefüllt. Füllgrad immer > 0,85
	Merkmal ist für die Quelle gefüllt. Niedriger bzw. wechselnder Füllgrad, siehe Merkmalsbeschreibung und Häufigkeitsauszählung
	Merkmal ist für diese Quelle nicht gefüllt. Füllgrad immer < 0,05

Beispiel: Das Merkmal Tagesentgelt bzw. täglicher Leistungssatz ist nur für BeH- und LeH-Sätze gefüllt. Die Sätze der anderen Datenquellen enthalten in dieser Variablen den Missing-Wert „.n“. Eine weitere Besonderheit liegt darin, dass einige Merkmale je nach Datenquelle unterschiedliche Inhalte haben. So enthält der Erwerbsstatus für BeH-Sätze die Personengruppe des Meldeverfahrens, für LeH-Sätze die Leistungsart, für LHG-Sätze den Status SGB II, für ASU- bzw. XASU-Sätze den Arbeitsuche-Status und für MTH die Maßnahmenkategorie. Nicht bei jedem Merkmal sind diese Unterschiede direkt aus dem Variablennamen ersichtlich.

**Tabelle 4 Merkmalsliste mit Füllgrad**

Merkmalsliste	BHP Basis	Seite	BeH	LeH	LHG	ASU	XASU	MTH
<b>Identifikatoren</b>		<b>42</b>						
Systemfreie Personennummer (persnr)		42						
Systemfreie Betriebsnummer (betnr)	*	43						
<b>Generierte technische Merkmale</b>		<b>44</b>						
Satzzähler pro Konto (spell)		44						
Quelle des Satzes (quelle)		44						
Jahr (jahr)	*	44						

<b>Merkmalsliste</b>	<b>BHP Basis</b>	<b>Seite</b>	<b>Beh</b>	<b>LeH</b>	<b>LHG</b>	<b>ASU</b>	<b>XASU</b>	<b>MTH</b>
<b>Gültigkeitszeitraum</b>		<b>45</b>						
Beginndatum Originalbeobachtung (begorig)		45						
Enddatum Originalbeobachtung (endorig)		45						
Beginndatum der Episode (begepi)		46						
Enddatum der Episode (endepe)		46						
<b>Informationen zur Person</b>		<b>46</b>						
Geschlecht (frau)		46						
Geburtsjahr (gebjahr)		47						
Staatsangehörigkeit (nation)		47						
Staatsangehörigkeit vergrößert (nation_gr)		47						
Familienstand (famst)		48						
Kinderzahl (kind)		48						
Ausbildung (ausbildung)		49						
Schulbildung (schule)		50						
<b>Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche</b>		<b>51</b>						
Tagesentgelt/ täglicher Leistungssatz (tentgelt)		51						
Beruf – ausgeübte/letzte Tätigkeit (KldB 1988) (beruf)		52						
Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_3)		53						
Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_4)		54						

<b>Merkmalsliste</b>	<b>BHP Basis</b>	<b>Seite</b>	<b>BeH</b>	<b>LeH</b>	<b>LHG</b>	<b>ASU</b>	<b>XASU</b>	<b>MTH</b>
Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau)		55	■	■		■	■	■
Teilzeit (teilzeit)		55	■					
Erwerbsstatus (erwstat)		56	■	■	■	■	■	■
Gleitzone (gleitz)		58	■					
Leiharbeit (leih)		58	■					
Befristung (befrist)		59	■					
Abmeldegrund / Abgabegrund / Beendigungsgrund (grund)		59	■	■	■	■	■	
Erwerbsstatus vor Arbeitsuche (estatvor)		60				■	■	
Status nach Arbeitsuche (estatnach)		61				■		
Profillage (profil)		61				■		■
Art der Kündigung der letzten Tätigkeit (art_kuend)		62				■		■
Arbeitszeit des Stellengesuchs (arbeitszeit)		62				■		■
Restanspruch/geplante Dauer (restanspruch)		62		■				■
Trägerart (traeger)		63			■	■	■	■
Beginndatum der Arbeitslosigkeit (alo_beg)		63	■	■	■	■	■	■
Dauer der Arbeitslosigkeit (alo_dau)		64	■	■	■	■	■	■
<b>Ortsangaben</b>		<b>64</b>						
Wohnort - Kreis (wo_kreis)		64	■	■	■	■	■	■
Wohnort - Bundesland (wo_bula)		65	■	■	■	■	■	■

<b>Merkmalsliste</b>	<b>BHP Basis</b>	<b>Seite</b>	<b>BeH</b>	<b>LeH</b>	<b>LHG</b>	<b>ASU</b>	<b>XASU</b>	<b>MTH</b>
Wohnort - Arbeitsagentur (wo_aa)		65						
Wohnort - Regionaldirektion (wo_rd)		66						
<b>Betriebsmerkmale</b>		<b>66</b>						
WZ73 3-Steller (w73_3)	*	66						
WZ93 5-Steller (w93_5)	*	67						
WZ93 3-Steller (w93_3)	*	67						
WZ03 5-Steller (w03_5)	*	68						
WZ03 3-Steller (w03_3)	*	68						
WZ08 5-Steller (w08_5)	*	69						
WZ08 3-Steller (w08_3)	*	69						
W73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen)	*	70						
Art der Vervollständigung w73_3 (group_w73_3)	*	70						
W93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen)	*	70						
Art der Vervollständigung w93_3 (group_w93_3)	*	71						
Jahr des 1. Auftretens der Betriebsnummer (grd_jahr)	*	71						
Erstes Auftreten Betriebsnummer (grd_dat)	*	72						
Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr)	*	72						
Letztes Auftreten Betriebsnummer (lzt_dat)	*	73						
Anzahl Beschäftigte gesamt (az_ges)	*	73						

<b>Merkmalsliste</b>	<b>BHP Basis</b>	<b>Seite</b>	<b>BeH</b>	<b>LeH</b>	<b>LHG</b>	<b>ASU</b>	<b>XASU</b>	<b>MTH</b>
Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz)	*	73						
Anzahl geringfügig Beschäftigte (az_gf)	*	74						
Mittelwert imputiertes Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)	*	74						
Arbeitsort Kreis (ao_kreis)	*	74						
Arbeitsort Bundesland (ao_bula)	*	75						

## 1.5 Mengengerüst

Tabelle 5 Mengengerüst

<b>Fallzahlen</b>	<b>vor Episodensplitting</b>	<b>nach Episodensplitting</b>
BeH	30.431	43.913
LeH	4.198	7.864
LHG	7.056	19.111
ASU	15.317	30.605
XASU	691	1.583
MTH	3.096	5.843
Sätze insgesamt	60.789	108.919
Personen		2.053

## 2 Datenquellen

### 2.1 IAB-SOEP Migrationsstichprobe

Das Ziel der IAB-SOEP-Migrationsstichprobe ist eine Verbesserung der Datengrundlage über Personen mit Migrationshintergrund in Deutschland und ihre Haushaltsmitglieder. Die Abgrenzung des Migrationshintergrundes zielt auf jüngere Migrationsbewegungen. Der Hauptfokus liegt hierbei auf den seit Mitte der 1990er Jahre zugewanderten Personen sowie auf Kindern von Einwanderern (als zweite Generation von Zuwanderern). Die Stichprobe liefert Detailinformationen zu deren familiärer Situation, sozialer Herkunft, Migrations- und Integrationserfahrung, Persönlichkeitsmerkmalen und Wertvorstellungen. Die Fokussierung der Untersuchungsperspektive auf die Integration von Zuwanderern erfordert ein als Längsschnitt angelegtes Untersuchungsdesign; die Teilnehmer der IAB-SOEP-Migrationsstichprobe werden

deswegen in jährlichem Turnus im Rahmen des Sozio-Oekonomischen Panels (SOEP) befragt.

## 2.2 Administrative Daten aus den IEB

Die administrativen Personendaten werden aus den Integrierten Erwerbsbiografien (IEB) des IAB gezogen und werden entsprechend der vom FDZ erstellten Stichprobe der integrierten Erwerbsbiographien (SIAB) aufbereitet. Diese vereinen Daten aus fünf verschiedenen Datenquellen, die jeweils Informationen aus unterschiedlichen EDV-Verfahren enthalten können. Zusätzlich werden einige ergänzende Merkmale aus diesen Datenquellen in die administrativen Personendaten aufgenommen, die nicht Teil der IEB sind. Abbildung 2 veranschaulicht die Datenflüsse, die zur SIAB führen und stellt deren Beziehung zu anderen Datenprodukten des FDZ dar.

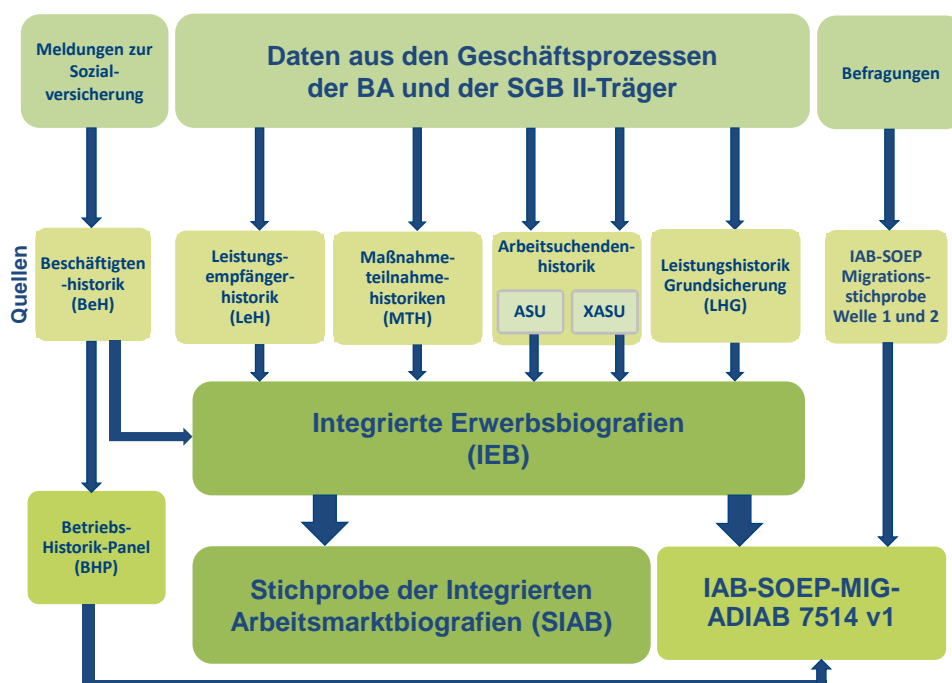


Abbildung 2 Quellen der IAB-SOEP-MIG-ADIAB 7514 v1

## 2.3 Beschäftigtenhistorik (BeH)

Die Quelle für Informationen über Beschäftigung ist die Beschäftigtenhistorik (BeH) des IAB. Die Grundlage der Daten ist das mit Wirkung vom 1. Januar 1973 eingeführte (und seit dem 1. Januar 1991 auf Ostdeutschland ausgedehnte) integrierte Meldeverfahren zur Kranken-, Renten- und Arbeitslosenversicherung, das durch das Kürzel DEÜV (früher DEVO/DÜVO) bezeichnet ist (vgl. für weitere Details: Bender et al., 1996, S. 4ff.; Wermter/Cramer, 1988). Es

verlangt von den Arbeitgebern, mindestens einmal jährlich Meldungen für alle sozialversicherungspflichtig beschäftigten Arbeitnehmer an die zuständigen Sozialversicherungsträger abzugeben. Die BeH erfasst alle Arbeiter und Angestellten sowie alle Auszubildenden, soweit sie nicht von der Sozialversicherungspflicht befreit sind. Dies bedeutet, dass verbeamtete und selbstständige Tätigkeiten sowie Studierende<sup>2</sup> (vgl. Cramer, 1985) grundsätzlich nicht in der BeH erfasst werden. Seit der Änderung des Meldeverfahrens zum 1. Januar 1999 werden auch geringfügig Beschäftigte und mithelfende Familienangehörige erfasst (enthalten erst ab 1. April 1999). Die Daten werden von den Krankenkassen aufgenommen, von der Bundesagentur für Arbeit gesammelt und aufbereitet und anschließend vom IAB in dessen Historikdatei integriert.

Die administrativen Personendaten werden durch Betriebsdaten (Basisbetriebsdaten und BHP-Erweiterungsmodule) ergänzt. Diese werden dem Betriebs-Historik-Panel (BHP) entnommen, das ebenso auf der BeH basiert.

Bei der Verknüpfung von Personen- mit Betriebsdaten muss berücksichtigt werden, dass die Merkmale in der Basisbetriebsdatei sowie den BHP-Erweiterungsmodulen zum 30.6. eines Jahres aggregiert werden. Die Betriebsmerkmale sind deshalb, im Unterschied zu den Personenangaben, keine zeitraumbezogenen Daten, sondern nur am 30.6. jedes Jahres exakt gültig.<sup>3</sup>

Die Verknüpfung der Betriebs- und Personendatei erfolgt über die programmspezifischen Kommandos der zur Datenaufbereitung und -analyse genutzten Softwarepakete. In Stata können die Daten beispielsweise unter Verwendung des „merge“ - Befehls und der Jahresangabe sowie der Betriebsnummer in Verbindung mit der entsprechenden Pfadangabe verknüpft werden (siehe Kasten 1).

```
use IAB-SOEP-MIG-ADIAB_7514_v1.dta
gen jahr = year(begepi)
sort betnr jahr
merge m:1 betnr jahr using IAB-SOEP-MIG-ADIAB_7514_v1_bhp_basis_v1.dta
```

#### **Kasten 1: Beispielcode für Stata 14**

## **2.4 Leistungsempfängerhistorik (LeH)**

Die Leistungsempfängerhistorik (LeH) erfasst Zeiträume, in denen Personen Lohnersatzleistungen der Bundesagentur für Arbeit beziehen (Rechtskreis SGB III). Die Leistungen bein-

---

<sup>2</sup> Studierende können dennoch in der BeH erscheinen wenn sie parallel zum Studium z.B. geringfügig beschäftigt waren.

<sup>3</sup> Extrembeispiel: Es liegen Beschäftigungsmeldungen vom 1.1. bis zum 30.5.2006 vor, der Betrieb meldet jedoch Konkurs an, weshalb alle Beschäftigten entlassen werden. Dann gibt es im BHP zu diesem Betrieb für 2006 keine Informationen.



halten Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld, also keine Leistungen nach dem Rechtskreis SGB II (z.B. Arbeitslosengeld II). Da der Leistungsanspruch von der Erfüllung gesetzlicher Voraussetzungen abhängt, werden Arbeitslosigkeitsperioden, in denen die Voraussetzungen nicht vorliegen (z.B. fehlende Bedürftigkeit im Falle der Arbeitslosenhilfe oder Nichterfüllung der Anwartschaftszeit beim Arbeitslosengeld) nicht in der Leistungsempfängerhistorik berichtet. Der linke Datenrand der LeH ist der 1. Januar 1975.

## **2.5 Leistungshistorik Grundsicherung (LHG)**

Die Leistungshistorik Grundsicherung (LHG) enthält Informationen zu erwerbsfähigen Leistungsberechtigten, ihren Bedarfsgemeinschaftsmitgliedern nach § 7 SGB II und besonderen Personen im Umfeld einer Bedarfsgemeinschaft. Eine Verknüpfung von Individuen mit SGB-II-Leistungsbezug auf Ebene der Bedarfsgemeinschaften ist in der SIAB jedoch nicht möglich. Leistungsbezug nach SGB II umfasst sowohl Leistungen zur Sicherung des Lebensunterhalts (z.B. ALG II), wie auch Zuschläge zum ALG II oder auch zusätzliche Leistungen. Die LHG enthält jedoch keine Angaben zu den jeweiligen Leistungssätzen. Da die Ermittlung der Höhe des Leistungsbezugs in der Grundsicherung nicht auf individueller Ebene erfolgt, sondern auf der Ebene von Bedarfsgemeinschaften, ist die Zuweisung eines individuellen Leistungssatzes problematisch und erfolgt in der SIAB nicht.

Im Gegensatz zu den Leistungen aus dem Rechtskreis des SGB III ist die Bundesagentur für Arbeit nicht alleiniger Träger der Leistungen. Es wird demnach unterschieden nach den drei möglichen Trägerarten:

- Arbeitsgemeinschaften (ARGE, bis 2010)/gemeinsame Einrichtungen (ab 2011) aus gemeinsamer Aufgabenwahrnehmung von BA und Kommune,
- Getrennte Trägerschaft/getrennte Aufgabenwahrnehmung (bis 2011) – hier besteht eine Aufgabenteilung zwischen BA und Kommune<sup>4</sup>,
- Zugelassene kommunale Träger, die nach der anfänglichen Experimentierklausel des § 6a auch als optierende Kommunen oder Optionskommunen bezeichnet werden – hier übernimmt die Kommune alle Aufgaben aus dem Bereich des SGB II.

---

<sup>4</sup> Die Kommune übernimmt die Kosten der Unterkunft und Heizung (§ 22 SGB II) sowie die abweichend zu erbringenden Leistungen (§ 23 (3) SGB II) und die flankierenden Leistungen nach §16 (2) Satz 2 Nr. 1 - 4 SGB II. Von der BA werden dagegen Regelleistungen, Sozialversicherungsbeiträge und Eingliederungsleistungen (SGB III und SGB II) sowie spezifische Leistungen ohne die oben genannten flankierenden Leistungen erbracht.

Die Daten der LHG stammen aus verschiedenen Meldeverfahren. In der Regel findet das IT-Verfahren A2LL bis 2010 bei ARGEn und ab 2011 bei gemeinsamen Einrichtungen Verwendung<sup>5</sup>. Zugelassene kommunale Träger verwenden verschiedene eigene IT-Verfahren und übermitteln ihre Daten mittels des Standards XSozial-BA-SGB II an die Bundesagentur für Arbeit. Bei Kommunen mit getrennter Trägerschaft kommen beide Verfahren zum Einsatz. Die unterschiedlichen Datenstandards haben Auswirkungen auf Umfang und Qualität der gelieferten Daten.

Der linke Datenrand der LHG ist der 1. Januar 2005. Die Datenquelle ist jedoch bis Anfang 2007 lückenhaft (siehe Kapitel 4).

## **2.6 Arbeitssuchendehistorik (ASU/XASU)**

Daten über Arbeitssuchende werden in der Arbeitssuchendehistorik (ASU und XASU) gespeichert. Die Quelle ASU enthält Informationen zu Arbeitssuchenden, welche bei Arbeitsagenturen verzeichnet sind, ab 2005 auch solche, die bei Arbeitsgemeinschaften und getrennten Trägerschaften gemeldet sind. In der Quelle XASU sind ab 2005 die Daten von arbeitssuchenden Personen im ALG-II-Bezug von zugelassenen kommunalen Trägern enthalten, welche nach dem XSozial-Standard konsolidiert gemeldet werden.

## **2.7 Maßnahmeteilnahmehistoriken (MTH)**

Die Maßnahmeteilnahmehistoriken (MTH) enthalten Informationen, die verschiedenen Rechtskreisen zugeordnet werden können. Zum einen sind Maßnahmen und Maßnahmeteilnahmen der aktiven Arbeitsmarktförderung nach dem SGB III enthalten. Zum anderen beinhalten die MTH Maßnahmen aus dem Rechtskreis SGB II, wenn diese in Verfahren der BA erfasst werden. Dies bedeutet insbesondere, dass in den MTH keine Maßnahmen der zugelassenen kommunalen Träger geführt werden, da diese über den abweichenden Standard XSozial gemeldet werden. Informationen von diesen Trägern werden aus einer Reihe datentechnischer Probleme nicht in die IEB aufgenommen. Der linke Datenrand der MTH ist der 1. Januar 2000.

# **3 Datenaufbereitung und Stichprobenziehung**

## **3.1 Bereinigungen und Aufbereitungen**

Bevor die Daten aus den in Kapitel 2 genannten Quellen zur IEB zusammengefügt werden, durchlaufen sie quellenspezifische Bereinigungsverfahren (siehe Folgeabschnitte). Die Gesamt-IEB erfahren die folgenden Bereinigungen:

---

<sup>5</sup> 2014 wurde A2LL sukzessive durch ALLEGRO als neues IT-Verfahren für die Grundsicherung im SGB II in gemeinsamen Einrichtungen ersetzt.

- Sätze, bei denen das Alter an ihrem Ende kleiner 13 oder zu Beginn größer 75 ist, werden gelöscht.
- Datensätze, bei denen das Enddatum kleiner als das Beginndatum ist, werden gelöscht.
- Widersprüchliche Angaben zu Geschlecht oder Geburtsdatum innerhalb eines Kontos werden bereinigt
- Datensätze ohne Angabe von Geburtsdatum und Geschlecht nach Bereinigung werden gelöscht.

Weitergehende Bereinigungen, wie sie z.B. in der IAB-Beschäftigtenstichprobe vorgenommen wurden (Ergänzen von vermutlich fehlenden Meldungen, Streikbereinigungen, etc.), erfolgen nicht.

### **Beschäftigtenhistorik (BeH)**

- Um einen im Zeitablauf möglichst konstanten Personenkreis zu erfassen, werden einige Personengruppen, deren Daten nicht für den gesamten Beobachtungszeitraum zur Verfügung stehen, ausgeschlossen. Ab dem Meldejahr 2011 stammen die Daten der BeH aus neu konzipierten Quelldaten. Dadurch wurde eine Reihe von Personengruppen reaktiviert bzw. neu eingeführt, da sie von der BA-Statistik als sozialversicherungspflichtig eingestuft werden. Damit sind fortan die Personengruppen 101 - 107, 111 - 114, 118, 119, 120, 140, 141, 142, 143, 149, 201 und 203 - 205 enthalten sowie die beiden Gruppen 109 und 209, die geringfügig Beschäftigte kennzeichnen. Nicht enthalten sind unter anderem kurzfristig Beschäftigte, d. h. Personengruppen 110, 202 und 210.
- Die Personengruppen 123, 124 und 127 wurden neu eingeführt.
- Aus datenschutzrechtlichen Gründen werden die Personengruppen 107, 111, 113, 114, 127 und 204 zur Personengruppe "Sonstige Erwerbstätige" (599) zusammengefasst.
- Ab dem Meldejahr 2012 wurden Auszubildende mit den Personengruppen 121 und 122 neu aufgenommen.
- Sätze mit einem Entgelt in Höhe von Null bzw. mit fehlendem Entgelt sowie der Ausprägung 101 beim Merkmal Erwerbsstatus und dem Wert 50 beim Abgabegrund (Jahresmeldung) werden nicht in die IEB übernommen.
- Geschlecht und Geburtsdatum werden aus dem Data Warehouse (DWH) der Statistik der BA übernommen. Dort liegen diese Angaben quellenübergreifend harmonisiert vor.
- Der Gebietsstand für Arbeitsort und Wohnort entspricht dem 31.12.2014.

### **Leistungsempfängerhistorik (LeH)**

- Sätze ohne gültiges Beginndatum wurden ausgeschlossen.
- Sätze mit einem Enddatum, das vor dem Beginndatum liegt, wurden ausgeschlossen.

- Liegt das Enddatum bei Bezug von Arbeitslosenhilfe einen Tag vor dem Beginndatum, so wurde das Enddatum um ein Jahr hochgesetzt und der Satz nicht gelöscht.
- Zwischen 2004 und 2006 wurde das zugrunde liegende Meldeverfahren umgestellt. Auftretende Überschneidungen zwischen dem Alt- und Neuverfahren wurden bereinigt.
- Sätze mit leerem oder ungültigem Enddatum wurden ausgeschlossen, da hier nicht davon ausgegangen werden kann, dass überhaupt eine Zahlung geleistet worden ist.
- In der SIAB wird eine Gebietsstandsberreinigung analog zur BeH durchgeführt.

### **Leistungshistorik Grundsicherung (LHG)**

- Sätze ohne BA-Kundennummer werden gelöscht.
- Konten ohne gültiges Geburtsdatum werden gelöscht.
- Stornierte Datensätze werden nicht genutzt.
- Es sind nur Sätze erwerbsfähiger Personen und von Personen über 65 Jahren enthalten.
- Es sind jeweils überschneidungsfreie Zeiträume der Leistungsberechtigung einer Person in einer bestimmten Bedarfsgemeinschaft abgebildet. Aus folgenden administrativen Gründen beginnt ein neuer Satz:
  - an gesetzlich festgelegten, für strukturelle Änderungen der Bedarfsgemeinschaft relevanten Geburtstagen (14, 15, 18 und 25) sowie dem individuell gültigen Renteneintrittsalter der Mitglieder einer Bedarfsgemeinschaft (siehe Abschnitt 0),
  - bei Veränderungen in der Zusammensetzung einer Bedarfsgemeinschaft (z.B. durch Zu-/Abgänge),
  - bei Merkmalsänderungen des Bedarfsgemeinschafts-Kunden sowie
  - bei Beginn und Ende eines Sanktionssachverhalts für Sätze ab dem 1. April 2006. Allerdings ist zu beachten, dass anhand der Daten weder die Dauer, die Art noch der Zeitpunkt der Verhängung bzw. des Beginns der Sanktion festgestellt werden kann. Grund hierfür ist das Fehlen eines entsprechenden Merkmals oder einer Ausprägung, welches den Beginn, die Art oder die Dauer der Sanktion anzeigt.
- Aus obigem Grund sind alle personenbezogenen Merkmale, die für die Quelle LHG gefüllt sind, für den gesamten Zeitraum des Satzes gültig.
- Doppelmeldungen aufgrund der Gebietsreformen 2009/2011 und Trägerneuorganisation 2012 wurden soweit möglich bereinigt.
- In der SIAB wird eine Gebietsstandsberreinigung analog zur BeH durchgeführt.

### **Arbeitsuchendehistorik ASU und XASU**

- Sätze, deren Enddatum vor dem Beginndatum liegt, wurden nicht in die ASU übernommen.
- Es erfolgt keine Konsolidierung der ASU-Sätze einer einzelnen Person. Es kann daher vorkommen, dass Überschneidungen zwischen ASU-Sätzen existieren.

- Die personenbezogenen Merkmale, die für die Quellen (X)ASU gefüllt sind, beziehen sich immer auf den Beginn des Satzes.
- Ein neuer ASU-Satz wird angelegt, sobald ein Statuswechsel (z.B. von arbeitsuchend zu arbeitslos) erfolgt. Dies gilt auch, wenn sich die Trägerart (Agentur für Arbeit, ARGE, gemeinsame Einrichtung, zugelassener kommunaler Träger, getrennte Aufgabenwahrnehmung) ändert. In der Datenbasis der ASU gibt es nur die Unterscheidung zwischen Sätzen mit Status "arbeitslos", "arbeitsuchend" sowie seit 2006 „ratsuchend“ und „ohne Status“. In den IEB wird jedoch ein zusätzlicher Status „krank“ unterschieden. Der Erwerbsstatus "krank" wird IEB-Sätzen zugeordnet, wenn in der ASU-Datenbasis
  - ein Vorgängersatz mit Status „arbeitslos“ existiert, lückenlos anschließt und den Abgangsgrund "krank" hat und
  - ein Nachfolgesatz mit Status „arbeitslos“ existiert und ebenfalls lückenlos anschließt
  - und der Satz selbst nicht den Status "arbeitslos" sondern den Status "nicht arbeitslos arbeitsuchend" hat.
- Im Gegensatz zur Quelle ASU werden in der XASU nur die Zustände
  - „nicht arbeitslos, aber arbeitsuchend“ und
  - „arbeitsuchend und gleichzeitig arbeitslos“ unterschieden.
- Krankheitsepisoden werden in der XASU im Gegensatz zur ASU bei der Bildung des Erwerbsstatus nicht berücksichtigt, da in den XASU-Daten Informationen zu Krankheiten nicht erschlossen sind. Bei der Berechnung der Arbeitslosigkeitsdauer unter Einschluss von XASU-Sätzen können daher krankheitsbedingte Lücken nicht identifiziert werden.
- In der XASU werden überschneidungsfreie Zeiträume bezogen auf eine einzelne Person abgebildet. Bei Änderungen folgender Merkmale wird für die XASU jeweils ein neuer Datensatz gebildet:
  - Wechsel des Arbeitsuche-Status
  - Wechsel der Verfügbarkeit
  - Wechsel des SGB-II-Trägers
  - Wechsel des Wohnortes
- Eine Gebietsstandsberreinigung wird analog zur BeH durchgeführt.

### **Maßnahmeteilnahmehistoriken (MTH)**

- Sätze mit einem Enddatum, das vor dem Beginndatum liegt, wurden ausgeschlossen.
- Sätze, die mehr als ein Jahr nach Maßnahmenende erstellt wurden, werden gelöscht, falls ein anderer Satz existiert, der innerhalb eines Jahres nach Abschluss der Maßnahme erstellt wurde.

- Es wird lediglich der aktuellste Datensatz zu einer einzelnen Maßnahmeteilnahme verwendet.
- In die IEB werden nur Maßnahmeteilnahmen aufgenommen, die als „tatsächlich stattgefunden“ klassifiziert sind. Teilnahmen, die nicht oder noch nicht stattgefunden haben, werden gelöscht. Teilnahmen werden auch dann als nicht stattgefunden gewertet, wenn während der Teilnahme an einer Maßnahme ein Löschdatum gesetzt wird.
- Bestimmte Maßnahmenarten werden nicht aufgenommen. Hierzu zählen Unterstützung der Beratung und Vermittlung, Mobilitätshilfen und reine Reha-Maßnahmen.

### **SGB-II-Anonymisierung**

Um das Deanonymisierungsrisiko zu verringern, wird in der SIAB grundsätzlich nur das Geburtsjahr ausgewiesen. In den Quellen LHG und (X)ASU besteht allerdings die Gefahr, dass durch die zeitliche Struktur der Beobachtungen dennoch das exakte Geburtsdatum sichtbar werden könnte. Sätze könnten systematisch an bestimmten Geburtstagen oder dem Renteneintritt bzw. am Tag davor enden oder aber an diesem Tag neu beginnen.

Um eine indirekte Bestimmung des exakten Geburtstags zu verhindern, wird daher das folgende Verfahren angewendet. Sätze, die am 18., 25. oder 65. Geburtstag geteilt wurden, werden zu einem einzelnen Satz zusammengefügt. Hierdurch geht außer dem exakten Geburtstag keine weitere Information verloren. Bei Sätzen, die am 15. Geburtstag oder dem Renteneintritt beginnen bzw. am Renteneintrittstag enden wird das Beginndatum (Enddatum) des Satzes auf den Beginn (das Ende) des jeweiligen Quartals gesetzt. Bei dieser Korrektur kommt es zu einer Verzerrung der Satzdauer. Korrigierte Sätze sind in der Variable "Erwerbsstatus" gekennzeichnet.

### **3.2 Episodensplitting**

Die administrativen Personendaten liegen „gesplittet“ vor<sup>6</sup>. Falls sich innerhalb eines Kontos Sätze überlappen, werden diese durch künstliche Sätze mit neuen Datumsangaben derart ersetzt, dass vollständig parallele oder überschneidungsfreie Zeiträume entstehen (Episodensplitting). Die Anzahl der Sätze erhöht sich dadurch (siehe Abbildung 3).

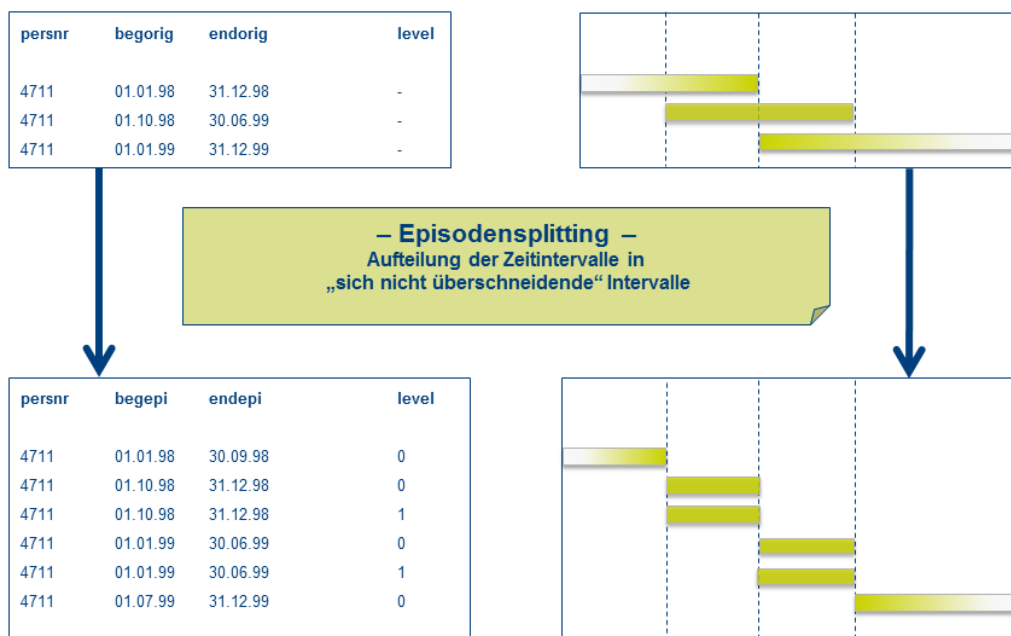
Die ursprünglichen Datumsvariablen Beginn und Ende des Originalsatzes (begorig und endorig) bleiben erhalten, die Variablen Beginn und Ende der Episode (begepi und endepi) kennzeichnen Anfang und Ende der gesplitteten Episoden. Ob Sätze gesplittet wurden, kann an einem Vergleich des ursprünglichen Zeitraums (begorig und endorig) mit dem Episodenzeitraum (begepi und endepi) festgestellt werden. Um die Originaldaten ohne Episodensplitting

---

<sup>6</sup> Durch das Record Linkage sind z.T. mehrere Konten einem Befragten zugeordnet. Bei diesen Personen, die mit Hilfe der Variable match\_az identifiziert werden können, sind die Daten für jedes gefundene Konto einzeln gesplittet.

wieder herzustellen bzw. die durch das Episodensplitting künstlich erzeugten Datensätze zu löschen, selektiert man alle Sätze, bei denen der Beginn des Originalsatzes gleich dem Beginn der gesplitteten Episode ist (`begepi == begorig`).

**Abbildung 3 Episodensplitting**



Es ist ratsam, die durch das Splitting erzeugten, vollständig parallelen Sätze einheitlich zu sortieren. Die früher in der SIAB enthaltenen Variablen Satzzähler pro Episode (`level2`) und Satzzähler pro Episode und Quelle (`level1`) können bei Bedarf mit folgenden Stata-Befehlen selbst erzeugt werden:

```
bysort persnr begepi quelle (spell): gen byte level1 = _n-1
```

```
bysort persnr begepi (spell): gen byte level2 = _n-1
```

### 3.3 Stichprobenziehung

Die IAB-SOEP Migrationsstichprobe ist aus allen Personen in den IEB Daten gezogen worden, deren IEB-Konto frühestens seit 1995 besteht. Es wurden Cluster mit je 2.500 Personen mit Hilfe von geokodierten Daten des FDZ gebildet, welche dann insgesamt 6.725 geografische Stichprobeneinheiten (PSUs) ergeben. Im nächsten Schritt wurden alle Personen mit Migrationshintergrund anhand der Nationalität und des Onomastik Verfahrens identifiziert. Das Onomastik Verfahren wurde nur in einer Teilstichprobe von 600 PSUs durchgeführt. Die Auswahl der PSUs folgte auf Grundlage geschichteter Zufallsstichproben (*stratified random sampling*).

Anschließend wurden alle Personen ohne Migrationshintergrund aus den Daten entfernt. Insgesamt wurden 250 aus den 600 PSUs für die Felderhebung berücksichtigt. In den ausgewählten 250 PSUs besteht die Bruttostichprobe aus 80 Adressen (Haushalte). Bei der Stichprobenziehung wurden entfernungsabhängige und simulierte random walk Verfahren implementiert. Durch ein sogenanntes Screening-Interview mit den Ankerpersonen wurden die Haushalte identifiziert, die die Kriterien des Stichprobendesigns erfüllen.

In Welle 2 wurden Haushalte aus der ersten Welle und ausgegründete Haushalte von befragten Personen aus der ersten Welle befragt, ferner wurden neue Haushaltsmitglieder in bestehenden und ausgegründeten Haushalten aufgenommen.

Die Stichprobe der Personen, die mit den administrativen Daten verknüpft werden können, ergibt sich aus der Bereitschaft der Befragten, einer Verknüpfung ihrer Befragungsdaten mit den administrativen Daten zuzustimmen.

### 3.4 Fehlende Werte

In den Daten gilt die folgende Codierung von fehlenden Werten:

Bezeichnung	Wert	Beschreibung
Keine (valide) Angabe	.z	Nicht systematisch fehlender Wert. Das Merkmal ist prinzipiell für die Quelle gefüllt, jedoch ist für den betrachteten Satz ein Wert nicht vorhanden bzw. nicht sinnvoll zu interpretieren.
Fehlt systematisch	.n	Ein Merkmal ist grundsätzlich für eine Quelle (weiße Felder der Merkmalsübersicht im Abschnitt 1.4) oder für einen gewissen Zeitraum nicht gefüllt.

## 4 Datenverknüpfung

Ausgangspunkt der Verknüpfung der Befragungs- und Prozessdaten sind die Befragten der IAB-SOEP-Migrationsstichprobe. Diese wurden im Rahmen des Interviews um Erlaubnis gebeten, ihre Befragungsdaten für Forschungszwecke mit den über sie am IAB vorhandenen administrativen Daten zu verknüpfen. Der Fragebogenausschnitt mit der Zustimmungserklärung lautet:

*„Ich bin damit einverstanden, dass Daten, die beim Institut für Arbeitsmarkt- und Berufsforschung (IAB) in Nürnberg vorliegen, mit in die Auswertung einbezogen und ausschließlich mit meinen Befragungsdaten der Studie LEBEN IN DEUTSCHLAND zusammengeführt und anonym gespeichert werden. Die Auswertung erfolgt, ohne dass ein Bezug zu meiner Person*



*möglich ist. Ich wurde darüber informiert, dass alle datenschutzrechtlichen Bestimmungen eingehalten werden. Ich behalte mir vor, mein Einverständnis jederzeit wieder zurückzuziehen.“*

Die Befragten mussten obigen Text unterschreiben, um Teil der Stichprobe für die Verknüpfung zu werden. Für das Vorgehen bei der Verknüpfung lassen sich die Befragten zwei Gruppen einteilen. Ankerpersonen sind diejenigen, die durch die Ziehung aus den Integrierten Erwerbsbiographien in die Stichprobe aufgenommen wurden und somit einen Indikator haben, mit dem sie wieder den IEB zugespielt werden können. Familienmitglieder der Ankerpersonen wurden ebenfalls befragt, haben aber keinen Indikator, mit dem die Daten den IEB zugespielt werden können. Diese Befragten müssen in den IEB mit Hilfe von Verknüpfungsverfahren gesucht werden.

In Welle 1 und 2 der IAB-SOEP Migrationsstichprobe wurden insgesamt 5.239 Personen mindestens einmal befragt, rund 75 Prozent wurden nach ihrem Einverständnis zur Verknüpfung gefragt. Davon gaben 2.235 ihre Zustimmung zur Verknüpfung (43 % aller Befragten, 58 % der Gruppe, die nach ihrem Einverständnis gefragt wurde). Für 2.053 Personen konnten die Befragungsdaten einem Konto (zum Teil zwei Konten) in den administrativen Daten zugeordnet werden.

Nach Abschluss der Befragung wurde versucht, zuspielbereite Befragungsteilnehmer zu verknüpfen. Generierte technische Merkmale im Datensatz kennzeichnen für jede Befragte und jeden Befragten, mit welcher Verknüpfungsmethode und mit welchem Maß an Sicherheit eine erfolgreiche Verknüpfung zustande gekommen ist. Damit kann jeder Nutzer nach eigenem Ermessen entscheiden, welche der Fälle in den eigenen Analysen Verwendung finden sollen.

#### **4.1 Bereinigungen und Standardisierung personenbezogener Merkmale**

Für die Suche nach den Familienmitgliedern wurden aus allen verfügbaren Adressdatenquellen folgende personenbezogene Merkmale herangezogen:

- Vor- und Nachname
- Geschlecht
- Adresse
- Geburtsdatum

Die Merkmale wurden vor der Verknüpfung Bereinigungsverfahren unterzogen, um eine möglichst hohe Übereinstimmung zwischen beiden Datenquellen zu erreichen.

Konkret wurden Umlaute und Sonderzeichen durch ihre jeweiligen ASCII-konformen Äquivalente ersetzt, alle Buchstaben in Großbuchstaben umgewandelt sowie überflüssige Leerzeichen und Piktationszeichen entfernt. Mehrere Vor- bzw. Nachnamen wurden in einem Feld

ohne Leerzeichen zusammengefasst. Neben diesen grundlegenden Bereinigungen werden variablenspezifische Routinen zur Standardisierung der Daten durchgeführt. So wird die Hausnummer aus einem kombinierten Straße /Hausnummernfeld extrahiert und alle übrigen Adresszusätze (z.B. Stockwerk, Appartement-Nr.) gelöscht. Bei Straßennamen werden häufig vorkommende Muster und Abkürzungen standardisiert. Bei Ortsnamen werden Anhänge auf eine einheitliche Form gebracht (z.B. bei Ortschaften mit zusätzlicher Information wie Fluss oder Region) bzw. entfernt (z.B. bei Berlin-Kreuzberg).

#### **4.2 Linkage der Ankerpersonen**

Zunächst wurden die Ankerpersonen mit den IEB-Daten verknüpft. Da die Daten über die Ankerpersonen aufgrund der Art der Ziehung bereits die Identifikatoren aus den IEB-Daten enthalten, können diese direkt zusammengespielt werden. Durch die Umschlüsselung zwischen IEB-Versionen kann es passieren, dass auch bei Ankerpersonen mehrere IEB Personennummern vorhanden sind. Ferner ist eine Qualitätssicherung der Verknüpfung auch bei Ankerpersonen sinnvoll, da die Möglichkeit besteht, dass der Interviewer die falsche Person als Ankerperson des Haushalts bestimmt hat.

#### **4.3 Exaktes Record Linkage**

Für die Familienmitglieder wurde zunächst ein exakter Abgleich der Felder zwischen den Einträgen der beiden Datenquellen für die Familienmitglieder durchgeführt.

Im ersten exakten Schritt war die vollständige Übereinstimmung in allen zur Verfügung stehenden linkage-relevanten Feldern ausschlaggebend, also Vor- und Nachname, Geburtsdatum, Geschlecht, Straße, Hausnummer, Postleitzahl sowie Ort. In den nächsten beiden Schritten wurde jeweils bei weiterhin nicht verknüpften Personen eine Abweichung einmal im Feld Postleitzahl und einmal im Feld Hausnummer toleriert, während die jeweils übrigen Felder weiterhin exakt übereinstimmen mussten. Im generierten technischen Merkmal `match_typ` wird ersichtlich, in welchem Schritt die Verknüpfung erfolgte. Es wurden nur Merkmalskombinationen berücksichtigt, bei denen das Risiko, zwei unterschiedliche Personen zu verknüpfen (sog. falsch-positive Matches), sehr gering war. Insgesamt wurde im Rahmen des exakten Record Linkage mit folgenden Merkmalskombinationen auf Übereinstimmung zwischen den beiden Datenquellen geprüft:

1. Name, Vorname, Geburtsdatum, Geschlecht, PLZ, Straße, Hausnummer, Ort
2. Name, Vorname, Geburtsdatum, Geschlecht, Straße, Hausnummer, Ort
3. Name, Vorname, Geburtsdatum, Geschlecht, PLZ, Straße, Ort

#### **4.4 Probabilistisches Record Linkage**

Für Personen, für die durch keinen der bisherigen Schritte eine Entsprechung in den administrativen Daten ermittelt werden konnte, wurden anschließend fehlertolerante Matching-Verfahren verwendet (vgl. Herzog et al., 2007). Dabei wurde für jedes abzugleichende Variablenpaar aus beiden Datenquellen ein Ähnlichkeitsmaß berechnet und diese in einem zusammenfassenden Qualitätsindex aufsummiert. Die Berechnungen werden mit der Software Merge Tool-Box (MTB, Version 0.742, vgl. Schnell et al. (2005)) durchgeführt. Bei der Berechnung des Ähnlichkeitsmaßes wurde ein probabilistisches Matching angewandt, bei dem rechnerisch berücksichtigt wird, dass die Wahrscheinlichkeit einer Übereinstimmung von Merkmalen von deren Varianz in der Grundgesamtheit abhängig ist. Hierbei werden so genannte m- und u-Parameter berechnet, die angeben, mit welcher Wahrscheinlichkeit die Angaben zu einer Person zwischen zwei verschiedenen Quellen variieren bzw. wie wahrscheinlich das Auftreten eines identischen Merkmals bei unterschiedlichen Personen ist. Beispielsweise deutet eine Übereinstimmung bei der Variable Geschlecht weniger auf einen tatsächlichen Match hin als auf eine Übereinstimmung beim Nachnamen und wird deshalb niedriger gewichtet. Die variablenspezifischen m- und u-Parameter, die dem probabilistischen Matching zugrunde liegen, wurden im Rahmen der Verknüpfung der 1. Welle von PASS ermittelt (vgl. Bachteler, 2008). Es werden alle zur Verfügung stehenden Merkmale zwischen den beiden Datenquellen verglichen und ein additives Ähnlichkeitsmaß berechnet (match\_qual). Je höher der Wert dieses Index, desto größer die Gesamtähnlichkeit der beiden Zeilen und desto wahrscheinlicher liegt eine tatsächliche Übereinstimmung vor.

Die Adresspaare wurden nach dem Ähnlichkeitsmaß sortiert und einem Sichtvergleich unterzogen. Die dadurch ermittelten Verknüpfungen wurden als probabilistische Matches klassifiziert und in den Datensatz aufgenommen.

## **5 Datenqualität und Probleme**

### **5.1 IAB-SOEP Migrationsstichprobe**

Non-Response, Gewichtung und andere Aspekte der Datenqualität finden Sie im SOEP Survey-Paper Nr. 216 unter [http://panel.gsoep.de/soep-docs/surveypapers/diw\\_ssp0216.pdf](http://panel.gsoep.de/soep-docs/surveypapers/diw_ssp0216.pdf).

### **5.2 Gesamte IEB**

Die IEB beinhalten Erwerbsbiografien. Allerdings ist in den administrativen Daten nicht jede Art von Erwerbstätigkeit erfasst. Bei entsprechenden Lebensverläufen sind bestimmte Personen überhaupt nicht in den IEB abgebildet.

Für Auswertungszwecke ist es oft relevant, Lücken in den abgebildeten Biografien zu erkennen (z.B. für Kontrollgruppenbildung, Lebensverlaufsanalysen etc.). Die im Folgenden aufgeführten Lücken sind definiert als Zeiträume nach Ende der Schulausbildung, für die kein Satz in den IEB enthalten ist. Diese Lücken können unterschieden werden in

- Lücken, für die keinerlei Information vorliegt und
- Lücken, für die anhand des Merkmals Abgabegrund/Beendigungsgrund/SGB-II-Einstellungsgrund/Abmeldegrund des direkten Vorgängersatzes eine Information vorliegen kann (sofern ein entsprechender Satz vorhanden ist).

Diese Lücken wurden identifiziert anhand der Ausprägungen der Merkmale Abgabegrund/Beendigungsgrund/SGB-II-Einstellungsgrund/Abmeldegrund und Erwerbsstatus in den verschiedenen Quellen. Die Auflistung erhebt keinen Anspruch auf Vollständigkeit.

**Tabelle 6 Biographische Lücken und Identifikationsmöglichkeiten**

<b>Biografische Lücke</b>	<b>Information über Lücke, ggf. identifizierbar anhand der Ausprägung des Merkmals „grund“ im Vorgängersatz der Quelle</b>
Beamte, Berufssoldaten, Richter, Angestellte von Körperschaften oder Stiftungen des öffentlichen Rechts	XASU
Ungeförderte Selbstständige	LeH, ASU
Studierende, Personen in schulischer Weiterbildung	LeH, LHG, ASU, XASU
Personen, die länger als 6 Wochen krank / arbeitsunfähig sind (Krankheit während Arbeitslosigkeit ist aber unter bestimmten Bedingungen in der Quelle ASU abgebildet, vgl. Abschnitt 0)	BeH, LeH, ASU
Personen in Altersrente ohne Beschäftigung wenn nicht Mitglied in einer BG	LeH, LHG, ASU
Personen in Mutterschutz / Elternzeit	XASU
Bezieher von Vorruhestandsgeld	LeH, ASU
Hausgewerbetreibende	
Kurzfristig Beschäftigte	ASU
Personen in Einrichtungen der Jugendhilfe, Berufsbildungswerken, anerkannten Werkstätten oder ähnlichen Einrichtungen für behinderte Menschen	ASU
Teilnehmende an Leistungen zur Teilhabe am Arbeitsleben (Rehabilitanden)	ASU
(Nebenerwerbs-)Landwirte	
Pflegepersonen i.S. von § 19 SBG XI	

Grundwehrdienstleistende	BeH, LeH, LHG, ASU, XASU
Wehrübungsleistende	BeH, LeH, LHG, ASU, XASU
Zivildienstleistende	BeH, LeH, LHG, ASU, XASU
Ableistende eines freiwilligen sozialen bzw. ökologischen Jahres anstelle des Zivildienstes	
Sonstige Personen, die nicht bei der Rentenversicherung oder der BA registriert sind (z.B. Sabbatjahr, Finanzierung aus privatem Vermögen oder Rentenbezügen, Emigration, Beschäftigung im Ausland, ehrenamtliche Tätigkeit, etc.)	BeH, LeH, ASU
Streikende bei Streikdauer über einem Monat	LeH
Sozialhilfe-Bezieher (vor Einführung des SGB II 2005), Sozialgeld-Bezieher (im Sinne des SGB II)	
Ausgleichsempfänger nach dem FELEG (Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit)	

### 5.3 Beschäftigtenhistorik (BeH)

- Die Einführung des neuen Tätigkeitsschlüssels in 2011 hat zu einer Reihe von Problemen geführt. So kam es während des Übergangszeitraums, der Arbeitgebern im Meldeverfahren zur Sozialversicherung eingeräumt wurde,<sup>7</sup> zu einer temporären Zunahme fehlender Angaben. Analysen der Statistik der BA (Bertat et al., 2013, S. 10) zeigen, dass die neuen bzw. umgestellten Merkmale „ausgeübte Tätigkeit“, „Arbeitszeit“ und „Ausbildungsabschluss“ nach der Umstellung zu etwa 20 bis 30 Prozent keine Angaben enthielten. Eine deutliche Verbesserung stellt sich im ersten Halbjahr 2013 ein. Um die Qualität der Arbeitszeitangabe im Übergangszeitraum zu verbessern, haben Ludsteck/Thomsen (2016) ein Imputationsverfahren entwickelt, das die fehlenden durch imputierte Werte ersetzt. Die imputierten Daten werden in der SIAB 7514 bereits verwendet.
- Aufgrund der Einführung des Meldeverfahrens in den neuen Bundesländern kann erst ab 1993 von einer hinreichenden Vollständigkeit der Meldungen ostdeutscher Betriebe ausgegangen werden. Aus demselben Grund weisen sehr viele Sätze des Jahrs 1991 für einige Merkmale (u.a. Erwerbsstatus) fehlende Werte auf.
- Der Anstieg der Zahl der BeH-Sätze ab 1999 folgt aus der Einführung der Meldepflicht für geringfügig Beschäftigte ab dem 1. April 1999.
- Es lässt sich insbesondere 1999 ein deutlicher Anstieg bei den Meldungen zu Teilzeitbeschäftigung feststellen. Verursacht wird dieser Befund sowohl vom tatsächlich zu beobachtenden Anstieg der Teilzeitbeschäftigung als auch von der Tatsache, dass seit 1999 die Beschäftigtenmeldungen generell besser ausgefüllt werden.

<sup>7</sup> Die im Rahmen des Meldeverfahrens eingesetzten Prüfprogramme ließen bis einschließlich Mai 2012 leere Angaben beim Tätigkeitsschlüssel 2010 zu.

- Innerhalb des Meldeverfahrens ist eine gewisse Zeitverzögerung nicht vermeidbar. Zwar müssen Änderungen von Beschäftigungsverhältnissen sofort gemeldet und bestehende Beschäftigungsverhältnisse jährlich bis zum April (bis Jahresmeldung 2012) bzw. Mitte Februar (ab Jahresmeldung 2013) des folgenden Jahres bestätigt werden. Faktisch treffen Meldungen jedoch zum Teil um Jahre verspätet ein. Die Historikdatei des IAB wird aber nicht kontinuierlich fortgeschrieben, sondern in bestimmten Abständen aktualisiert. Dies geschieht unter Verwendung von Dateien mit den Beschäftigten-Meldungen jeweils eines Jahres, die 36, 18, 12 oder 6 Monate nach Ablauf des Berichtsjahres eingetroffen sind (frühestens im Juli 2015 z.B. kann die 18-Monats-Datei für das Jahr 2013 erstellt werden). Mehr als 3 Jahre verspätete Meldungen werden im IAB nicht berücksichtigt, so dass eine 36-Monats-Datei per definitionem einen Füllgrad von 100% aufweist.
- In der zugrundeliegenden Version der IEB liegt der Füllgrad der BeH-Meldungen zum Jahr 2012 bei knapp 100%, da für dieses Jahr nur das 30-Monatsmaterial verwendet werden konnte.<sup>8</sup> Für das Jahr 2013 wurde eine 18-Monats-Datei verwendet, für 2014 eine 6-Monats-Datei. Es ist daher davon auszugehen, dass die BeH-Meldungen für das Jahr 2013 leicht und für 2014 etwas stärker untererfasst sind. Die Auswertbarkeit der Daten auf Personenebene sollte dadurch aber nicht eingeschränkt sein. Allerdings häufen sich die fehlenden Meldungen in wenigen Betrieben. Dies bedeutet, dass in Einzelfällen die Betriebsangaben, z.B. die Betriebsgröße, grob falsch sind und sich in Folgeversionen stark verändern werden.
- Im Jahr 1984 erfolgte eine Änderung des Meldeverfahrens. Einmalig gezahltes Bruttoarbeitsentgelt wurde ab diesem Zeitpunkt als Teil des beitragspflichtigen Jahresentgelts gemeldet, was zu einer Erhöhung des durchschnittlichen Tagesentgelts führt. Insbesondere steigt ab diesem Jahr der Anteil der Entgelte oberhalb der Beitragsbemessungsgrenze stark an (vgl. Bender et al., 1996).
- Für die Jahre 1992 bis 2000 wurden in den IEB auffällige Einbrüche und Anstiege in der Anzahl der Meldungen festgestellt. Es lassen sich v.a. Einbrüche bei folgenden zehn Kreisen beobachten: Braunschweig (03101), Wolfsburg (03103), Emden (03402), Kassel (06633), Essen (05113), Neuss (05162), Erftkreis (05362), Hersfeld-Rotenburg (06632), Miltenberg (09676) und Kempten (Allgäu) (09763). Grund hierfür sind Meldeprobleme bei einzelnen bzw. mehreren Betrieben in diesen Regionen.

---

<sup>8</sup> Durch ein Redesign der Datengrundlage der BeH stand zum Zeitpunkt der Erstellung der BeH ausnahmsweise nur das 30- statt des 36-Monatsmaterials zur Verfügung. Analysen mit früheren Datenversionen haben jedoch gezeigt, dass i.d.R. bereits die 18-Monats-Datei einen Füllgrad von rund 99% aufweist. Demnach kommt es in den weiteren 18 Monaten nur noch zu extrem wenigen Nach- oder Korrekturmeldungen.

- Nennenswerte Einbrüche verzeichnen auch die Kreise Salzgitter (03102) und Hoyerswerda (14264).
- Bei den Meldungen zu Vollzeitbeschäftigung in den IEB sind besonders die Kreise Main-Taunus-Kreis (06436) und Alzey-Worms (07331) erwähnenswert. Hier sind überdurchschnittliche Anstiege zu beobachten. Als Grund sind ebenfalls Meldeprobleme bei einzelnen bzw. mehreren Betrieben in diesen Regionen anzuführen.
- In den Jahren 1996 bis 1998 kommen die Werte 841-844 (Ärzte und Apotheker) in der Variable Beruf – ausgeübte Tätigkeit (beruf) im Vergleich zu benachbarten Jahren sehr selten vor. Die Gründe dafür sind nicht bekannt.

#### **5.4 Leistungsempfängerhistorik (LeH)**

- Für die neuen Bundesländer werden die LeH-Sätze erst ab 1992 vollständig erfasst.
- Die Leistungsbezugsdaten wurden früher auf Magnetbändern gesichert. Aufgrund eines Schadens an einem Magnetband sind die Leistungsbezugsdaten bis einschließlich des Jahres 1980 nur teilweise enthalten. Daher ist für das vorliegende Datenprodukt auch davon auszugehen, dass die Informationen zum Leistungsbezug im genannten Zeitraum nur zum Teil enthalten sind.
- Durch einen internen Systemwechsel kommt es 2004 zu einem Bruch in der Erfassung von Sperr- und Säumniszeiten. Bis 1. Juli 2004 können Sperr und Säumniszeiten nur durch den Abmeldegrund im vorhergehenden LeH-Satz identifiziert werden. Danach existiert ein eigener Satz mit täglichem Leistungssatz = 0 für Versagens-, Entziehens- und Ruhezeiten.

#### **5.5 Leistungshistorik Grundsicherung (LHG)**

- In den Jahren 2005 und 2006 ist die LHG nicht vollständig gefüllt. Es kommt zu erheblichen trägerspezifischen Ausfällen. Es wird daher abgeraten, für diesen Zeitraum Analysen allein auf Basis der LHG zu machen.
- Längsschnittuntersuchungen für Personen sind mit Unsicherheiten behaftet, da Umzüge in und aus Gebieten von Trägern mit Lieferproblemen nicht von Änderungen des Status der Leistungsberechtigung unterschieden werden können.
- Auch ab 2007 kommt es hauptsächlich bei zugelassenen kommunalen Trägern immer wieder zu meist einmonatiger trägerspezifischer Untererfassung.
- Im Zusammenhang mit Trägerartwechseln kommt es zu Unter-/Übererfassungen:
  - Im Rahmen der Trägergebietsreform vom 1. Januar 2011 kommt es zu Untererfassungen im Gebiet der Arbeitsagenturen Dessau-Roßlau, Halberstadt, Halle und Sangerhausen.



- Doppelmeldungen aufgrund der Gebietsreformen 2009/2011 und der Trägerformwechsel zum 1. Januar 2012 werden in der IEB bereits soweit möglich bereinigt. Dennoch können Doppelmeldungen auftreten.
- Bei folgenden Jobcentern bestehen Unschärfen bei der Zuordnung von Leistungsfällen:
  - zwischen Emden und Norden zwischen September und Dezember 2009
  - zwischen Döbeln und Mittelsachsen von Oktober bis Dezember 2012
  - zwischen Tirschenreuth und Wunsiedel von November 2012 bis März 2013
- Manche Personen, für die eine LHG-Beobachtung vorliegt, sind vom Leistungsbezug nach SGB II ganz oder teilweise ausgeschlossen, da sie z. B. eine geförderte Ausbildung machen, Altersrente beziehen, in einer stationären Einrichtung untergebracht sind oder Versicherungsleistungen zur Vermeidung von Hilfebedürftigkeit bekommen. Dies betrifft pro Jahr ca. 3 bis 5 Prozent der Fälle. In XSozial kommt es zu trägerspezifischen Untererfassungen dieser Personengruppe. In der SIAB ist der Ausschluss von Leistungen nicht identifizierbar.

## **5.6 Arbeitssuchendehistorik (ASU/XASU)**

### **ASU**

- Die registrierten Zeiten der Arbeitssuche der Quelle ASU können ab 1997 als vollständig angesehen werden. Das Analysepotenzial der ASU-Sätze vor 1997 ist damit eingeschränkt.
- Für die Vermittlungskräfte ist es nicht möglich, die Zuordnung zum Rechtskreis immer unmittelbar zu erfassen, da vielfach erst zeitverzögert feststeht, welcher Träger aufgrund einer möglichen SGB-II-Leistungsberechtigung vorrangig zuständig ist. Es wird daher empfohlen, die Ausprägung des Merkmals „Trägerart“ in der ASU mit der Ausprägung in der LHG im gleichen Zeitraum abzugleichen. Aufgrund der Erfassungslücken bei LHG im Zeitraum 2005 bis 2006 ist dies nicht immer möglich.
- Für manche Personen, für die ab 2005 eine Optionskommune zuständig war, wurden von der BA parallel "künstliche" ASU-Sätze angelegt. Diese sind über estatvor (Wechsel zu einem zugelassenen kommunalen Träger) zu identifizieren.
- Von Mitte 2005 bis Mitte 2006 wurde in der BA das EDV-Verfahren coArb, aus dem die Arbeitssuchendaten stammen, durch das Verfahren VerBIS abgelöst. Ab Juli 2005 wurde coArb erstmals in der Agentur Wiesbaden als Pilotprojekt, ab Dezember 2005 dann sukzessive in mehreren Tranchen in allen Arbeitsagenturen durch VerBIS ersetzt. Viele der erfassten Merkmale wurden in den beiden Systemen in unterschiedlicher Differenziertheit und unterschiedlicher qualitativer Gewichtung abgefragt. Die Integration dieser Merkmale in die IEB gestaltet sich daher schwierig und ist nur durch die Anwendung eines speziellen Verfahrens (Mapping) möglich. Leider kann durch das Mapping keine vollständige Umsetzung der betroffenen Merkmale von coArb auf VerBIS erreicht werden, sodass sich bei



einigen Merkmalen die Ausprägungen „Keine Angabe“, „Sonstiges“ oder „Missing“ ungewöhnlich häufen. Außerdem können sich bei Häufigkeitsauszählungen auffällige Unterschiede ergeben, je nachdem ob die Ursprungsquelle der Daten coArb oder VerBIS war. Wichtige Einschränkungen des Analysepotentials werden in der jeweiligen Variablenbeschreibung angesprochen.

- Das bis Juni 2006 genutzte Verfahren coArb unterstützte ausschließlich die Vermittlung von Arbeitslosen und Arbeitsuchenden. Lückenhaft erfasst wurden auch Daten zu Personen, die sich lediglich beraten lassen wollten. Die Daten der Berufsberatung wurden in einem separaten System gehalten. Mit VerBIS erweiterten sich die Ausprägungen des Arbeitsuche-Status um „Ratsuchende“ und Personen „ohne Status“. Zu Letzteren gehören ALG-II-Leistungsberechtigte, die der Arbeitsvermittlung eingeschränkt zur Verfügung stehen. Ihre Erfassung in VerBIS gilt erst seit Januar 2008 als weitgehend vollständig.
- Ein Wechsel der Grundsicherungsträger oder des Wohnorts führen nicht zu einem neuen ASU-Satz, die Merkmalsausprägung jeweils zu Beginn einer Episode wird fortgeführt. Mit zunehmender Dauer des Spells wird also das Risiko größer, dass Träger oder Wohnort nicht mehr zutreffend sind.

## **XASU**

- Im Gegensatz zu den Arbeitsuchendendaten aus den ARGE n und den getrennten Träger-schaften zeigen sich nach dem 1.1.2005 für die zugelassenen kommunalen Träger zwischenzeitlich systematische Untererfassungen. Deshalb sollten Daten aus der Quelle XASU erst ab 2007 ausgewertet werden.
- Verschiedene Merkmale weisen für die XASU zum Teil nur einen sehr geringen Füllgrad auf. Hiervon betroffen sind unter anderem die Merkmale Schulabschluss, Abgabegrund sowie Erwerbsstatus vor Arbeitsuche. Die Füllgrade dieser Merkmale verbessern sich zwar im Zeitablauf, bleiben aber dennoch zum Teil unbefriedigend. Das Merkmal Beruf ist nahezu für den gesamten zur Verfügung stehenden Zeitraum in der XASU nicht gefüllt.
- Bei etlichen Trägern (Kreisen) erscheint der Anteil der gemeldeten Arbeitsuchenden an den gemeldeten erwerbsfähigen Hilfebedürftigen in den IEB durchgängig oder zeitweise un-plausibel hoch. Gründe dafür könnten eine möglicherweise fehlerhafte Bestimmung des Status „Nichtarbeitslos, arbeitssuchend“ oder eine technisch fehlerhafte Übermittlung mit hohem Anteil an Nachmeldungen bei diesen Trägern sein.
- Die träger- und zeitraumabhängige Plausibilität der XASU-Angaben sollte jeweils unter Berücksichtigung der Untersuchungsfrage vor der Verwendung geprüft werden.

## **5.7 Maßnahmeteilnahmehistoriken (MTH)**

- Die MTH sind für Maßnahmen mit einem Eintrittsdatum vor dem 1.1.2000 unvollständig.

- Zum 1.1.2005 kommt es zu einem Bruch, da mit Einführung des SGB II Maßnahmeteilnehmer auf verschiedene Träger aufgeteilt wurden (siehe Abschnitte 2.5 und 2.7).
- Die MTH enthält nur Meldungen, die in BA-Verfahren gebucht sind. Deren Nutzung nimmt in Arbeitsgemeinschaften/Getrennten Trägerschaften/getrennten Aufgabenwahrnehmungen zwischen 2005 und 2007 kontinuierlich zu. Ab März 2007 sind die Meldungen für diese Träger komplett.
- Maßnahmen, die über den Standard XSozial gemeldet werden, sind nicht in der MTH oder der SIAB enthalten.
- Durch die Neuorganisation von SGB-II-Trägern kann die Dokumentation von Maßnahmeteilnahmen in der MTH abrechnen oder neu beginnen, wenn es zu einer Umstellung des Meldeverfahrens kommt.<sup>9</sup>
- Bei Meldungen zum Überbrückungsgeld wird zum Teil die maximal zulässige Dauer von einem halben Jahr überschritten. Dies kann in den meisten Fällen durch eine Voreinstellung der Eingabemaske im Erfassungssystem erklärt werden.
- Die MTH wird um Bewerbermerkmale (z. B. Ausbildung) aus anderen BA-Verfahren ergänzt. Für diese Merkmale erfolgte im Jahr 2006 die Umstellung des operativen Verfahrens von coArb auf VerBIS. Es gelten hier die gleichen Qualitätseinschränkungen wie für die ASU.

## 6 Variablenbeschreibung

Auszählungen sowie Übersichten über die einzelnen Ausprägungen und Labels der Merkmale finden sich in den Arbeitshilfen zur IAB-SOEP-MIG ADIAB unter <http://fdz.iab.de>.

### 6.1 Identifikatoren

#### Systemfreie Personennummer (persnr)

Variablenlabel	Systemfreie Personennummer
Variablenname	persnr
Kategorie	Identifikatoren
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine

<sup>9</sup> Weitere Informationen zur Gebietsstruktur der SGB-II-Träger und relevanten Änderungen finden sich unter: <http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Regionale-Gliederungen/Gebietsstruktur-Traeger-Grundsicherung-Nav.html>

Detailbeschreibung	<p>Die systemfreie Personennummer zeigt an, welche Sätze zur gleichen Person gehören. Systemfrei bedeutet, dass von dieser Personennummer nicht auf Eigenschaften der Person oder Original-Identifikatoren geschlossen werden kann.</p> <p>Da es in den verschiedenen Datenquellen keinen einheitlichen Personen-Identifikator gibt, ist die Zuordnung der Informationen aus den unterschiedlichen Datenquellen (z.B. Beschäftigung und Leistung) zu Personen(-Konten) nicht immer eindeutig. In solchen Fällen kann es zu unplausiblen Erwerbshistoriken kommen.</p> <p>Die Bildung des quellenübergreifenden Personen-Identifikators basiert auf einer von der BA entwickelten Heuristik.</p>
--------------------	---

### Systemfreie Betriebsnummer (betnr)

Variablenlabel	Systemfreie Betriebsnummer
Variablenname	betnr
Kategorie	Identifikatoren
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Die systemfreie Betriebsnummer zeigt an, welche Sätze zum gleichen Betrieb gehören. Sie basiert auf der von der BA vergebenen Betriebsnummer, die durch eine systemfreie Nummer ersetzt wurde. (Weitere Informationen zur Betriebsnummernvergabe der BA finden sich in Bender et al., 1996, S. 15f. und S. 27-30.) Die Betriebsnummer kann zusammen mit der Jahresangabe benutzt werden, um Personen- und Betriebsinformationen zusammenzuspielen.</p> <p>Für die Betriebsnummer sollte Folgendes grundsätzlich beachtet werden:</p> <ul style="list-style-type: none"> <li>a) Besteht das Unternehmen aus nur einer Niederlassung oder hat das Unternehmen in einer Gemeinde nur eine Niederlassung, so ist diese Niederlassung der Betrieb und erhält eine Betriebsnummer.</li> <li>b) Hat das Unternehmen <b>mehrere</b> Niederlassungen in <b>einer</b> Gemeinde, so sind diese Betriebs- und Arbeitsstätten zu einem Betrieb unter einer Betriebsnummer zusammenzufassen, wenn sie derselben Wirtschaftsklasse angehören bzw. ist jede Niederlassung ein Betrieb und erhält je eine Betriebsnummer, wenn sie <b>nicht</b> derselben Wirtschaftsklasse angehören.</li> <li>c) Hat das Unternehmen <b>mehrere</b> Niederlassungen in <b>mehreren</b> Gemeinden, so ist jede dieser Niederlassungen ein Betrieb und erhält je eine Betriebsnummer.</li> </ul> <p>In diesem Zusammenhang sind auch folgende Definitionen im Sinne der Betriebsnummernvergabe im Rahmen des Meldeverfahrens zur Sozialversicherung zu beachten:</p> <ul style="list-style-type: none"> <li>a) Ein <b>Betrieb</b> ist eine regional und wirtschaftsfachlich abgegrenzte Einheit, in der Beschäftigte tätig sind und für die nach o.a. Grundsätzen eine Betriebsnummer zu vergeben wird.</li> </ul>

	<ul style="list-style-type: none"> <li>b) Eine <b>Arbeitsstätte</b> ist eine Einheit, in der Beschäftigte tätig sind und für die nach o.a. Grundsätzen eine Betriebsnummer nicht zu vergeben ist.</li> <li>c) Ein <b>Unternehmen</b> als Begriff fasst Betriebs- und Arbeitsstätten desselben Arbeitgebers zusammen.</li> <li>d) Ein <b>Arbeitgeber</b> ist jede natürliche oder juristische Person, für die mindestens ein sozialversicherungspflichtig bzw. geringfügig Beschäftigter tätig ist.</li> <li>e) Betrieb und Betriebsstätten sind synonyme Begriffe; Niederlassung ist dann ein synonyme Begriff für Filialen, Zweigstellen, Auslagerungen, Dependancen, Arbeitsstätten u.ä., wenn sie kein Betrieb ist.</li> </ul>
Qualitätshinweis	Die Betriebsnummer aus den IAB ist zum Teil nicht in der BeH enthalten. Bei diesen Meldungen handelt es sich in den meisten Fällen um Meldungen, die den Stichtag 30.6. nicht überschneiden. Des Weiteren werden keine Betriebe im BHP, die keinen Beschäftigten mit Entgelt>0 beschäftigen oder nur Beschäftigte mit erwstat=599 (sonstige Beschäftigte). Da Betriebsmerkmale (Arbeitsort, Wirtschaftszweig, Betriebsgröße usw.) über die Betriebsnummer zugespielt werden, fehlen sie in diesen Meldungen.

## 6.2 Generierte technische Merkmale

### Satzzähler pro Konto (spell)

Variablenlabel	Satzzähler pro Konto
Variablenname	Spell
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Der Satzzähler Konto nummeriert aufsteigend die Sätze einer Person. Das Merkmal wird beim Episodensplitting erzeugt und bezieht sich auf die gesplitteten Sätze. Mit dem Merkmal „Satzzähler Konto“ kann die ursprüngliche Sortierung auf einfache Art und Weise wiederhergestellt werden. Die Sortierung erfolgt grundsätzlich zunächst nach dem Beginndatum der Episode und dann nach der Quelle.

### Quelle des Satzes (quelle)

Variablenlabel	Quelle des Satzes
Variablenname	Quelle
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal gibt die Datenquelle an (siehe Kapitel 2).

### Jahr (jahr)

Variablenlabel	Jahr
----------------	------

Variablenname	Jahr
Kategorie	Generierte technische Merkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Dieses Merkmal ist nur in der Betriebsdatei enthalten. Es gibt das Gültigkeitsjahr der Betriebsinformationen zum Stichtag 30. Juni an. Mit Hilfe dieses Merkmals und der Betriebsnummer können Personen- und Betriebsdatei verknüpft werden.

### 6.3 Gültigkeitszeitraum

#### Beginndatum Originalbeobachtung (begorig)

Variablenlabel	Beginndatum Originalbeobachtung
Variablenname	Begorig
Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	<p>Das Beginndatum des Originalsatzes entspricht dem ursprünglichen Beginndatum der Meldung. Dies kann vom Beginndatum der betreffenden Datenzeile, dem sogenannten Episoden-Beginndatum, abweichen (vgl. auch die Erläuterungen zum Episodensplitting in Abschnitt 3.2). Da die Meldelogik unter bestimmten Umständen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben in diesen Fällen durch das in Abschnitt 0 beschriebene Anonymisierungsverfahren verändert.</p> <p><b>1) BeH</b> In BeH-Meldungen sind Beginn- und Endjahr wegen der Regeln des Meldeverfahrens immer identisch (Pflicht des Arbeitgebers zur Abgabe von Jahresmeldungen). Ein durchgehendes Beschäftigungsverhältnis kann also auf mehrere Meldungen verteilt sein.</p> <p><b>2) LHG, ASU, XASU</b> Bestimmte Änderungen führen zur Anlage eines neuen Datensatzes (siehe Abschnitte 0 und 0). begorig gibt das Beginndatum des neuen Zeitraums an.</p>

#### Enddatum Originalbeobachtung (endorig)

Variablenlabel	Enddatum Originalbeobachtung
Variablenname	Endorig
Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Datum
Hierarchie	Keine

Detailbeschreibung	<p>Das Enddatum des Originalsatzes entspricht dem ursprünglichen Enddatum der Meldung. Dies kann vom Enddatum der betreffenden Datenzeile, dem sogenannten Episoden-Enddatum, abweichen (vgl. auch die Erläuterungen zum Episodensplitting in Abschnitt 3.2). Da die Meldelogik unter bestimmten Umständen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben in diesen Fällen durch das in Abschnitt 0 beschriebene Anonymisierungsverfahren verändert.</p> <p><b>1) BeH</b> In BeH-Meldungen sind Beginn- und Endjahr wegen der Regeln des Meldeverfahrens immer identisch (Pflicht des Arbeitgebers zur Abgabe von Jahresmeldungen). Ein durchgehendes Beschäftigungsverhältnis kann also auf mehrere Meldungen verteilt sein.</p> <p><b>2) LHG, ASU, XASU</b> Bestimmte Änderungen führen zur Anlage eines neuen Datensatzes (siehe Abschnitte 0 und 0). endorig gibt das Enddatum des neuen Zeitraums an.</p>
--------------------	--

#### Beginndatum der Episode (begepi)

Variablenlabel	Beginndatum der Episode
Variablenname	Begepi
Kategorie	Generierter Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Beginndatum der Episode. Es ist immer gleich dem oder größer als das Beginndatum der Originalbeobachtung (vgl. auch die Erläuterungen zum Episodensplitting in Abschnitt 3.2).

#### Enddatum der Episode (endepe)

Variablenlabel	Enddatum der Episode
Variablenname	Endepe
Kategorie	Generierter Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Enddatum der Episode. Es ist immer gleich dem oder kleiner als das Enddatum der Originalbeobachtung (vgl. auch die Erläuterungen zum Episodensplitting in Abschnitt 3.2).

### 6.4 Informationen zur Person

#### Geschlecht (frau)

Variablenlabel	Geschlecht
Variablenname	Frau

Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Geschlechter-Dummy (0 - Mann, 1 - Frau). Die Geschlechtsangabe ist innerhalb eines Personenkontos konstant.

### Geburtsjahr (gebjahr)

Variablenlabel	Geburtsjahr
Variablenname	Gebjahr
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Geburtsjahr ist innerhalb eines Personenkontos konstant.
Qualitätshinweis	In den Originaldaten kann es vorkommen, dass das Geburtsjahr zwischen den Quellen wechselt. Dies wird bei der Datenaufbereitung bereinigt. Hierbei kommt Angaben aus dem Leistungsbezug die höchste Priorität zu.

### Staatsangehörigkeit (nation)

Variablenlabel	Staatsangehörigkeit
Variablenname	nation
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal enthält als Ausprägungen die Staatenschlüssel des Statistischen Bundesamtes.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt.

### Staatsangehörigkeit vergrößert (nation\_gr)

Variablenlabel	Staatsangehörigkeit vergrößert
Variablenname	nation_gr
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal enthält als Ausprägungen die zu Gruppen zusammengefassten Staatenschlüssel des Statistischen Bundesamtes.

## Familienstand (famst)

Variablenlabel	Familienstand
Variablenname	famst
Kategorie	Persönliche Merkmale
Herkunft	LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal beschreibt den Familienstand. In der Quelle LeH umfasst das Merkmal nur zwei Ausprägungen (0 - nicht verheiratet, 1 - verheiratet), während in den Quellen LHG/ASU/XASU/MTH zwischen sechs Ausprägungen unterschieden wird. Ein Abgleich der Informationen aus den Quellen wurde nicht vorgenommen.

## Kinderzahl (kind)

Variablenlabel	Kinderzahl
Variablenname	kind
Kategorie	Persönliche Merkmale
Herkunft	LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Je nach Datenquelle hat das Merkmal eine unterschiedliche Bedeutung.</p> <p><b>1) LeH</b>  In der LeH gibt die Variable die Anzahl der Kinder unter 16 Jahren zum Zeitpunkt der Antragstellung an. Es wird nur unterschieden, ob die Anzahl an Kindern gleich Null oder mindestens eins ist. Folgende Ausprägungen kommen daher vor:  0       Keine Kinder  100     Ein oder mehrere Kinder  Das Merkmal wird bei Änderungen der Leistungsart oder neuen Bewilligungen nicht aktualisiert, sondern erst, wenn ein erneuter Leistungsfall auftritt. Dadurch ist die Datenqualität eingeschränkt.</p> <p><b>2) ASU, MTH</b>  Hier entspricht die Ausprägung des Merkmals der Anzahl an Kindern unter 15. Bis zum 30.6.2006 konnten nur bis zu neun Kinder erfasst werden. Der Wert 0 kommt nicht vor. Vor dem 30.6.2006 wurde der Wert Null auf missing umgeschlüsselt, da nicht klar ist, ob Null inhaltlich als "Keine Kinder" oder als "Feld nicht bearbeitet" zu interpretieren ist. Nach dem 30.6.2006 ist das Merkmal nur gefüllt, wenn Kinder vorhanden sind.</p> <p><b>3) XASU, LHG</b>  Hier entspricht die Ausprägung des Merkmals der Anzahl an Kindern unter 15 in der Bedarfsgemeinschaft. Bei der Quelle LHG gilt der Wert für den gesamten Originalzeitraum.</p>



## Ausbildung (ausbildung)

Variablenlabel	Ausbildung
Variablenname	ausbildung
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Zu beachten ist, dass das Merkmal je nach Quelle eine unterschiedliche Bedeutung hat</p> <p><b>1) BeH</b>  Das Merkmal enthält für BeH-Sätze die von den Arbeitgebern im Rahmen des Meldeverfahrens anzugebende Berufsausbildung. Folgende Ausprägungen kommen vor:</p> <ul style="list-style-type: none"> <li>1 ohne abgeschlossene Berufsausbildung</li> <li>2 betriebliche/außerbetriebliche Ausbildung</li> <li>11 Fachhochschulabschluss</li> <li>12 Hochschulabschluss</li> </ul> <p>Bei Meldungen nach dem neuen Tätigkeitsschlüssel ist es nicht mehr möglich, FH-Absolventen wie bisher eindeutig zu identifizieren, da der neue Tätigkeitsschlüssel die separate Erfassung dieses Bildungsabschlusses nicht länger erlaubt. Sie werden der Kategorie 12 zugeordnet.</p> <p>„Zusammen mit einem Betriebswechsel treten häufig auch Änderungen im Ausbildungsstatus auf. Das liegt daran, dass beim neuen Betrieb die Meldedaten neu zusammengestellt werden. Wenn ein Beschäftigter z.B. durch berufsbegleitende Weiterbildung einen höheren Abschluss erlangt hat, dann wird dieser Statuswechsel wahrscheinlich erst von einem neuen Beschäftigungsbetrieb erfasst. Es ist in der Regel davon auszugehen, dass bei längeren Beschäftigungszeiten innerhalb eines Betriebs die einmal erhobenen personenbezogenen Daten fortgeschrieben werden.“ (Meinken/Koch, 2004, S. 63)</p> <p>Eine Methode, fehlende Werte oder inkonsistente zeitliche Verläufe der Bildungsangaben in der Vorgängerstichprobe IABS zu bereinigen, findet sich in Fitzenberger et al., 2006, sowie in Drews, 2006. Hierfür wird aber nur die Quelle BeH herangezogen, die in der IABS die einzige Datenquelle mit Informationen zur Bildung war.</p> <p><b>2) ASU, XASU, MTH</b>  Für diese Sätze ist die letzte abgeschlossene Berufsausbildung enthalten. Folgende Ausprägungen kommen vor:</p> <ul style="list-style-type: none"> <li>1 ohne abgeschlossene Berufsausbildung</li> <li>2 betriebliche/außerbetriebliche Ausbildung</li> <li>3 Berufsfachschule</li> </ul>

	<p>4 Fachschule  5 Fachhochschule  6 Universität  7 in Deutschland nicht anerkannte Berufsausbildung  8 in Deutschland nicht anerkannter Hochschulabschluss</p> <p>Wobei die Ausprägungen 7 und 8 nur für die Quelle MTH gelten.</p>
Qualitätshinweis	<p>Die Zahl der fehlenden Werte steigt im Zeitverlauf kontinuierlich an und erreicht am rechten Datenrand über 40%. Fehlende Werte treten besonders häufig bei folgenden Personengruppen auf: geringfügig beschäftigte Personen, Personen in Teilzeit, ausländische Beschäftigte und ostdeutsche Arbeitnehmer. Hintergrund ist, dass das Merkmal versicherungsrechtlich keine besondere Bedeutung hat (siehe Meinken/Koch, 2004, S. 63).</p> <p>Durch die Umstellung von coArb auf VerBIS kann in den Quellen ASU und MTH zwischen 2006 und 2008 nicht korrekt zwischen „ohne abgeschlossene Berufsausbildung“ und „Keine Angabe“ unterschieden werden. Ein fehlender Wert in diesem Zeitraum besagt also nicht unbedingt, dass keine (Information zur) Ausbildung vorliegt, sondern kann auch bedeuten, dass das einschlägige Generierungsverfahren nicht angewandt werden konnte. Der Füllgrad in der Quelle XASU ist generell niedrig.</p>

### Schulausbildung (schule)

Variablenlabel	Schulausbildung
Variablenname	schule
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal enthält den Schulabschluss. Je nach Quelle sind unterschiedliche Ausprägungen möglich.</p> <p><b>1) BeH</b>  Mit der Umstellung des Tätigkeitsschlüssels ändern sich die möglichen Ausprägungen. Davor gelten:</p> <p>5 Volks-/Hauptschule o./m. Abschluss, Mittlere Reife oder gleichwertige Schulbildung  8 Fachhochschulreife oder Abitur/Hochschulreife  9 Abitur/Hochschulreife</p> <p>Danach gelten:</p>

	<p>1 Kein Hauptschulabschluss  4 Hauptschulabschluss/Volksschulabschluss  6 Mittlere Reife  8 Fachhochschulreife oder Abitur/Hochschulreife</p> <p><b>2) ASU, XASU, MTH</b>  Für diese Quellen sind die folgenden Ausprägungen möglich:</p> <p>1 Kein Hauptschulabschluss  4 Hauptschulabschluss/Volksschulabschluss  6 Mittlere Reife  7 Fachhochschulreife  9 Abitur/Hochschulreife</p> <p>Sie sind zu Beginn der Arbeitssuche/Maßnahmeteilnahme gültig. In der Quelle XASU kann das Merkmal bei Ausbildungsplatzsuchenden auch den angestrebten Abschluss enthalten.</p>
Qualitätshinweis	Der Füllgrad nimmt in der BeH über die Jahre kontinuierlich ab und liegt in jüngeren Jahren unter 2/3. In der XASU nimmt er kontinuierlich zu und liegt seit 2012 über 2/3. In ASU und MTH ist der Füllgrad insgesamt gut.

## 6.5 Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

### Tagesentgelt/ täglicher Leistungssatz (tentgelt)

Variablenlabel	Tagesentgelt/täglicher Leistungssatz
Variablenname	tentgelt
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p><b>1) BeH</b></p> <p>In BeH-Sätzen weist das Merkmal jeweils das Bruttotagesentgelt eines Beschäftigten aus. Es wird berechnet aus dem vom Arbeitgeber gemeldeten Zeitraumentgelt sowie der Dauer des Zeitraums der (ungesplitteten) Originalmeldung in Kalendertagen. Das Tagesentgelt ist in Euro angegeben.</p> <p>Bis 1998 meldeten Arbeitgeber grundsätzlich nur das sozialversicherungspflichtige Arbeitsentgelt. Entgelte, die unterhalb der Geringfügigkeitsgrenze lagen, wurden nicht gemeldet. Entgelte, welche die Beitragsbemessungsgrenze in der gesetzlichen Rentenversicherung überschritten, werden nur bis zu dieser Grenze gemeldet. In der gesetzlichen Rentenversicherung gibt es zwei Beitragsbemessungsgrenzen. Die Grenze der Knappschaftlichen Rentenversicherung liegt meist höher als die Grenze der Arbeiter- und Angestelltenrentenversicherung. Jedoch ist eine Unterscheidung der beiden Versicherungsträger in den Daten nicht möglich.</p>

	<p>Seit der Einbeziehung der geringfügig Beschäftigten in das Meldeverfahren zum 1.4.1999 werden auch Entgelte unterhalb der Geringfügigkeitsgrenze erfasst. Die Beitragsbemessungsgrenze als obere Kapazitätsgrenze gilt jedoch weiterhin. In einigen Fällen übersteigen die gemeldeten Entgelte die Beitragsbemessungsgrenze allerdings trotzdem. Dies dürfte in der Regel auf Jahressonderzahlungen zurückzuführen sein, die der Arbeitgeber bei den Jahres-, Unterbrechungs- oder Abmeldungen zum normalen Entgelt hinzurechnen kann. Dabei ist es dann unerheblich, ob durch diese Addition die für den Meldezeitraum maßgebliche Beitragsbemessungsgrenze in der Rentenversicherung überschritten wird. Es könnte sich allerdings auch um fehlerhafte Angaben bei der Beschäftigungszeit handeln. (Die Entgelt-Informationen hingegen können auf Grund ihrer Versicherungsrelevanz als weniger fehleranfällig gesehen werden.)</p> <p>Die Geringfügigkeits- und Beitragsbemessungsgrenzen unterscheiden sich von Jahr zu Jahr und in West- und Ostdeutschland (maßgeblich ist jeweils der Standort des Betriebs). Eine Übersicht über diese Grenzen findet sich unter <a href="http://fdz.iab.de">http://fdz.iab.de</a>.</p> <p>Eine Angabe von 0 Euro ist auf "Unterbrechungsmeldungen" zurückzuführen. In diesen Zeiten besteht das Beschäftigungsverhältnis zwar rechtlich fort, es wird aber kein Entgelt gezahlt. Dies ist der Fall bei Krankheit nach Ende der Lohnfortzahlung, Mutterschaftszeiten und Sabbaticals.</p> <p>Das Tagesentgelt wird mit zwei Nachkommastellen ausgewiesen. Alle Werte größer als 0 und kleiner als 0,01 wurden auf 0,01 aufgerundet. Somit können die o.a. Unterbrechungsmeldungen mit der Bedingung Tagesentgelt = 0 identifiziert werden.</p> <p><b>2) LeH</b></p> <p>Bei LeH-Datensätzen enthält das Merkmal den täglichen Leistungssatz, jeweils umgerechnet in Euro. Es ist zu berücksichtigen, dass der tägliche Leistungssatz bei Datensätzen mit einem Original-Beginndatum, das vor dem 1.1.1998 liegt, für Werktage und bei Datensätzen mit einem Original-Beginndatum ab 1.1.1998 für Kalendertage gilt.</p> <p>Die Angabe von 0 Euro ist seit 1.1.2005 auf Sperrzeiten oder Leistungsunterbrechungen zurückzuführen. Ist für einen Satz mit Leistungssatz gleich 0 ein Beendigungsgrund angegeben, so handelt es sich um eine Unterbrechungsmeldung. Bei Sätzen, die eine Sperrzeit widerspiegeln, ist der Anspruch derselbe wie vor Beginn der Sperrzeit.</p>
--	--

**Beruf – ausgeübte/letzte Tätigkeit (KIdB 1988) (beruf)**

Variablenlabel	Beruf – ausgeübte/letzte Tätigkeit (KIdB 1988)
Variablenname	beruf
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine

Detailbeschreibung	<p><b>1) BeH</b> Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V).</p> <p>Der Arbeitgeber verschlüsselt die Tätigkeit der/des Beschäftigten hierbei gemäß der „Klassifizierung der Berufe 1988“ (Bundesanstalt für Arbeit, 1988), in der ca. 25.000 Berufsbezeichnungen zu finden sind. Die Berufsordnung besteht aus einer 3-stelligen Kennziffer und umfasst rund 330 Ausprägungen.</p> <p>Beschäftigungsmeldungen mit einem Enddatum nach dem 30.11.2011 werden nach dem neuen Tätigkeitsschlüssel 2010 (KldB2010) gemeldet. Diese Werte werden durch einen Schwerpunktumstieg auf die KldB1988 zurückgeschlüsselt. Dabei kommt es zu Unschärfen.</p> <p><b>2) ASU, XASU, MTH</b> Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1).</p>
Qualitätshinweis	<p>In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.</p> <p>In der XASU ist das Merkmal Beruf nahezu für den gesamten zur Verfügung stehenden Zeitraum nicht gefüllt.</p>

### Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010\_3)

Variablenlabel	Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010), 3-Steller
Variablenname	beruf2010_3
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p><b>1) BeH</b> Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V).</p> <p>Der Arbeitgeber verschlüsselt die Tätigkeit der/des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Die weniger detaillierte Berufsgruppe wird durch die ersten drei Stellen der Kennziffer erfasst. Das in der KldB2010 an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt.</p>

	<p>Beschäftigungsmeldungen mit einem Enddatum vor dem 30.11.2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet. Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten.</p> <p><b>2) LeH, ASU, XASU, MTH</b> Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1).</p>
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.

#### **Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010\_4)**

Variablenlabel	Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010), 4-Steller
Variablenname	beruf2010_4
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>1) BeH Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V). Der Arbeitgeber verschlüsselt die Tätigkeit der/des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Die weniger detaillierte Berufsuntergruppe wird durch die ersten vier Stellen der Kennziffer erfasst. Das in der KldB2010 an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt. Beschäftigungsmeldungen mit einem Enddatum vor dem 30.11.2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet. Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten.</p> <p>2) LeH, ASU, XASU, MTH Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1).</p>
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.

Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird als Berufsangabe gemäß KldB2010 lediglich die Berufsgruppe (beruf2010_3) bereitgestellt.
----------------	---

### Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau)

Variablenlabel	Anforderungsniveau - ausgeübte/letzte Tätigkeit (KldB 2010)
Variablenname	niveau
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>1) BeH Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V).</p> <p>Der Arbeitgeber verschlüsselt die Tätigkeit der/des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Das an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt.</p> <p>Beschäftigungsmeldungen mit einem Enddatum vor dem 30.11.2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet. Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten.</p> <p>2) LeH, ASU, XASU, MTH Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1).</p>
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.

### Teilzeit (teilzeit)

Variablenlabel	Teilzeit
Variablenname	teilzeit
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine



Detailbeschreibung	<p>Die berufliche Stellung eines Beschäftigten im Meldezeitraum wird vom Arbeitgeber im Rahmen der „Angaben zur Tätigkeit“ übermittelt.</p> <p>Das Merkmal „Berufsstellung“ unterscheidet zwischen Vollzeit- und Teilzeitbeschäftigten. Maßgeblich ist dabei das Verhältnis zwischen der arbeitsvertraglich vereinbarten und der betriebsüblichen Arbeitszeit. Für Teilzeitbeschäftigte wird nur erfasst, ob ihre Stundenzahl eine bestimmte Grenze übersteigt oder nicht. Diese Grenze lag bis 1978 bei 20 Wochenstunden, zwischen 1979 und 1987 bei 15 Wochenstunden und seit 1988 bei 18 Wochenstunden.</p> <p>Nur für Vollzeitbeschäftigte liefert die Variable tatsächlich Angaben zur beruflichen Stellung, wobei unter anderem zwischen vollzeitbeschäftigten Arbeitern und Angestellten sowie Auszubildenden unterschieden wird. Mit der Einführung des neuen Tätigkeitsschlüssels fiel diese Unterscheidung jedoch weg. Die Variable „teilzeit“ unterscheidet daher im kompletten Meldezeitraum lediglich zwischen Vollzeit- und Teilzeitbeschäftigung. Weitere Informationen zur Stellung im Beruf werden nicht verwendet.</p>
Qualitätshinweis	<p>In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Um dieses Problem abzumildern, wurde am IAB eine Imputation der Arbeitszeit für den fraglichen Zeitraum vorgenommen. Nähere Informationen zum Verfahren können Ludsteck/Thomsen (2016) entnommen werden.</p>

### Erwerbsstatus (erwstat)

Variablenlabel	Erwerbsstatus
Variablenname	erwstat
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal nimmt für jede Datenquelle unterschiedliche Werte mit unterschiedlichen Bedeutungen an.</p> <p><b>1) BeH</b></p> <p>Für BeH-Sätze entspricht das Merkmal Erwerbsstatus der im neuen Meldeverfahren (DEÜV) ab dem 1.1.1999 erfassten Personengruppe. Sie weist beitrags- oder leistungsrechtliche Besonderheiten des Beschäftigungsverhältnisses aus.</p> <p>Treffen für eine Beschäftigungsmeldung mehrere Schlüssel zu, ist vom meldenden Arbeitgeber der kleinste anzugeben. Mehrheitlich handelt sich dabei um sozialversicherungspflichtige Beschäftigungsverhältnisse ohne Besonderheiten, die unter Schlüsselnummer 101 geführt werden. Entsprechend ist es möglich, dass diese Beschäftigungen leicht überschätzt werden.</p> <p>Im Meldeverfahren ist geregelt, dass Wechsel im Erwerbsstatus - z.B. die Übernahme eines Auszubildenden nach dem Ende der Berufsausbildung im Ausbildungsbetrieb - durch eine erneute Meldung erfolgen müssen (vgl. Deutsche BKK, 2016, S. 35).</p>



Die Personengruppe kann in Beschäftigungsmeldungen, die sich auf die Jahre vor 1999 beziehen, aber erst 1999 oder später eingegangen sind, enthalten sein. Für Meldungen, die vor 1999 eingegangen sind, wird versucht, anhand bestimmter Regeln mit Hilfe der Merkmale „Ausbildung“, „Stellung im Beruf und Arbeitszeit“ und „Beruf“ sowie weiterer Informationen eine Zuordnung zu den Personengruppen durchzuführen. In vielen Fällen ist das jedoch nicht eindeutig möglich.

Seit dem 1.4.1999 werden im Meldeverfahren DEÜV auch geringfügig Beschäftigte erfasst. Diese Personengruppe lässt sich über die Ausprägungen 109 und 209 abgrenzen. Für geringfügig Beschäftigte konnten keine Daten vor dem Zeitpunkt der Einführung der Meldepflicht im Jahr 1999 zurückgeschrieben werden.

## **2) LeH**

Für LeH-Sätze ist im Merkmal Erwerbsstatus die gruppierte Leistungsart angegeben. Es kann also unterschieden werden, ob eine Person Arbeitslosengeld, Arbeitslosenhilfe oder Unterhaltsgeld bezieht oder ob Beiträge zur privaten Pflegeversicherung von der BA übernommen werden.

## **3) ASU/XASU**

Für ASU-Sätze zeigt das Merkmal Erwerbsstatus den Status der Arbeitssuche an. Als Ratsuchende werden Empfänger von Arbeitslosengeld (bzw. Arbeitslosengeld II), die Leistungen unter den erleichterten Voraussetzungen des § 428 SGB III (bzw. des § 65 Abs. 4 SGB II) beziehen, sowie Nichtleistungsempfänger ohne Vermittlungsbereitschaft i. S. des § 252 Abs. 8 SGB VI, wenn das 58. Lebensjahr vollendet wurde, geführt.

Der Status "ohne Status" (Statistik: „nicht gesetzt“) impliziert hauptsächlich Personen, denen nach §10 SGB II eine Aktivierung bzw. Vermittlung nicht zuzumuten ist. Ebenso werden Personen mit einer Arbeitsunfähigkeit von mehr als 42 Tagen, die weiterhin ALG II beziehen, unter diesem Status im System geführt.

In XASU-Sätzen enthält das Merkmal Erwerbsstatus bislang die Ausprägungen „nicht arbeitslos, aber arbeitsuchend“ sowie „arbeitsuchend und gleichzeitig arbeitslos“.

Da die Meldelogik in vielen Fällen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben durch das in Abschnitt 0 beschriebene Anonymisierungsverfahren verändert.

## **4) LHG**

Für LHG-Sätze zeigt das Merkmal Erwerbsstatus an, ob die Person minderjährig erwerbsfähig, volljährig erwerbsfähig oder nicht erwerbsfähig ab Altersrentengrenze geführt wird.

Da die Meldelogik in vielen Fällen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben durch das in Abschnitt 0 beschriebene Anonymisierungsverfahren verändert.

## **5) MTH**

	Für MTH-Sätze zeigt das Merkmal Erwerbsstatus die Maßnahmenart-Kategorie an. Dies ist die höchste Hierarchiestufe der Maßnahmenart-Systematiken der Bundesagentur für Arbeit.
Qualitätshinweis	In der LHG ist zu beobachten, dass es überdurchschnittlich viele 15-jährige und in geringerem Maß 16- und 17-jährige mit Hinweis auf fehlende Erwerbsfähigkeit gibt. 15- bzw. 16-jährige Hilfebedürftige der zugelassenen kommunalen Träger sind deshalb möglicherweise unterrepräsentiert, da „Nichterwerbsfähige“ nicht in die IEB übernommen werden.

### Gleitzone (gleitz)

Variablenlabel	Gleitzone
Variablenname	gleitz
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal ist erst ab 2003 und nur für BeH-Datensätze gefüllt. Es zeigt an, ob es sich bei der Beschäftigungsmeldung um eine Beschäftigung im Niedriglohnbereich innerhalb der sogenannten Gleitzone mit einem monatlichen Bruttoarbeitsentgelt von 400,01 bis 800,00 € handelt (sog. Midi-Job), für die der Arbeitnehmer nur einen reduzierten Beitrag zum Gesamtsozialversicherungsbeitrag leisten muss. Da Arbeitnehmer mit einem Entgelt in der Gleitzone freiwillig den „normalen“ Sozialversicherungsbeitrag zahlen können, sind nicht automatisch alle Beschäftigungen mit einem entsprechen Entgelt als Gleitzone gekennzeichnet. Die entsprechende gesetzliche Regelung besteht seit dem 1.4.2003.

### Leiharbeit (leih)

Variablenlabel	Leiharbeit
Variablenname	leih
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal gibt an, ob es sich bei der Beschäftigung der Person um ein Leiharbeitsverhältnis handelt. Das Merkmal entstammt dem Tätigkeitsschlüssel 2010 und ist nur für Meldungen mit einem Enddatum nach dem 30.11.2011 gefüllt.
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.

## Befristung (befrist)

Variablenlabel	Befristung
Variablenname	befrist
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	Das Merkmal gibt an, ob es sich bei der Beschäftigung der Person um ein befristetes oder unbefristetes Arbeitsverhältnis handelt. Das Merkmal entstammt dem Tätigkeitsschlüssel 2010 und ist nur für Meldungen mit einem Enddatum nach dem 30.11.2011 gefüllt.
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.

## Abmeldegrund / Abgabegrund / Beendigungsgrund (grund)

Variablenlabel	Abmeldegrund / Abgabegrund / Beendigungsgrund
Variablenname	grund
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, LHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p><b>1) BeH</b> In BeH-Meldungen kennzeichnet der Abgabegrund den Anlass, aus dem der Arbeitgeber den Sozialversicherungsträgern die betreffende Beschäftigungsmeldung übermittelt hat. Allerdings werden in die IEB nicht alle im Rahmen des Meldeverfahrens mögliche Abgabegründe integriert. So beinhalten die IEB nur Meldungen mit Entgeltangaben (also Jahres-, Unterbrechungs- und Abmeldungen), während Anmeldungen wegen fehlender Entgeltinformationen nicht enthalten sind. Ein Informationsverlust ist damit jedoch nicht verbunden, da die Angaben aus einer Anmeldung mit der darauf folgenden Jahres-, Unterbrechungs- oder Abmeldung erneut übermittelt werden.</p> <p>Die Abgabegründe für Beschäftigungsmeldungen sind nach den Regeln des seit 1.1.1999 geltenden Meldeverfahrens (gemäß DEÜV) kodiert.</p> <p><b>2) LeH</b> In LeH-Meldungen gibt das Merkmal den Grund für das Ende des Bezugs von Arbeitslosengeld, -hilfe oder Unterhaltsgeld an. Über die Gründe für den Beginn des Leistungsbezugs finden sich in der LeH keine Informationen, da die LeH mit den Meldungen der Arbeitsagenturen an die Krankenkassen über abgeschlossene Leistungsbezugsdauern gefüllt wird.</p> <p><b>3) LHG</b></p>

	<p>Bei LHG-Datensätzen ist der SGB-II-Einstellungsgrund enthalten und gibt den Grund für die Beendigung von laufenden Leistungen an. Der SGB-II-Beendigungsgrund <b>bezieht sich auf die Person</b>, nicht aber auf die Bedarfsgemeinschaft. Wird bei einem anderen als dem betrachteten Mitglied der Bedarfsgemeinschaft der SGB-II-Leistungsbezug beendet, werden neue Sätze für alle Mitglieder in einer Bedarfsgemeinschaft zu diesem Datum begonnen, der SGB-II-Beendigungs-/Einstellungsgrund ist jedoch nur für die betroffene Person gefüllt. Das Merkmal ist genau zum Ende des Originalsatzes gültig.</p> <p><b>4) ASU</b>  Bei ASU-Sätzen ist der Abmelde- bzw. Abgangsgrund enthalten. Im Fall eines Rechtskreiswechsels wird der Satz künstlich gesplittet und als Abmeldegrund "DS durch Splittung erzeugt" eingetragen. Zur korrekten Abbildung der Abgangsgründe ist außerdem das Merkmal „Status nach Arbeitssuche“ zu beachten.  Ab dem 26.4.2003 erfolgte eine Reduktion der Merkmalsausprägungen. Für Auswertungen über längere Zeiträume können die alten Werte auf die aktuell gültigen anhand folgender Tabelle umgeschlüsselt werden:</p> <table border="0"> <tr> <td><b>alt-&gt;neu</b></td> <td><b>alt-&gt;neu</b></td> <td><b>alt-&gt;neu</b></td> <td><b>alt-&gt;neu</b></td> </tr> <tr> <td>29-&gt;60</td> <td>36-&gt;61</td> <td>44-&gt;74</td> <td>51-&gt;74</td> </tr> <tr> <td>30-&gt;60</td> <td>37-&gt;66</td> <td>45-&gt;77</td> <td>52-&gt;76</td> </tr> <tr> <td>31-&gt;61</td> <td>38-&gt;66</td> <td>46-&gt;67</td> <td>53-&gt;68</td> </tr> <tr> <td>32-&gt;60</td> <td>39-&gt;71</td> <td>47-&gt;67</td> <td>54-&gt;78</td> </tr> <tr> <td>33-&gt;60</td> <td>40-&gt;69</td> <td>48-&gt;78</td> <td></td> </tr> <tr> <td>34-&gt;60</td> <td>42-&gt;65</td> <td>49-&gt;69</td> <td></td> </tr> <tr> <td>35-&gt;60</td> <td>43-&gt;70</td> <td>50-&gt;75</td> <td></td> </tr> </table> <p><b>5) XASU</b>  Bei XASU-Sätzen ist der Abmelde- bzw. Abgangsgrund enthalten. Im Fall eines Rechtskreiswechsels wird der Satz künstlich gesplittet und als Abmeldegrund "DS durch Splittung erzeugt" eingetragen.</p>	<b>alt-&gt;neu</b>	<b>alt-&gt;neu</b>	<b>alt-&gt;neu</b>	<b>alt-&gt;neu</b>	29->60	36->61	44->74	51->74	30->60	37->66	45->77	52->76	31->61	38->66	46->67	53->68	32->60	39->71	47->67	54->78	33->60	40->69	48->78		34->60	42->65	49->69		35->60	43->70	50->75	
<b>alt-&gt;neu</b>	<b>alt-&gt;neu</b>	<b>alt-&gt;neu</b>	<b>alt-&gt;neu</b>																														
29->60	36->61	44->74	51->74																														
30->60	37->66	45->77	52->76																														
31->61	38->66	46->67	53->68																														
32->60	39->71	47->67	54->78																														
33->60	40->69	48->78																															
34->60	42->65	49->69																															
35->60	43->70	50->75																															
Qualitätshinweis	Der Anteil valider Werte (Füllgrad) des Abgabegrundes ist in der Quelle LHG über alle Jahre sehr niedrig (< 20%). Auch in der XASU ist der Füllgrad gering, jedoch verbessert er sich im Zeitablauf.																																

#### Erwerbsstatus vor Arbeitssuche (estatvor)

Variablenlabel	Erwerbsstatus vor Arbeitssuche
Variablenname	estatvor
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitssuche
Herkunft	ASU, XASU
Datentyp	Numerisch
Hierarchie	keine

Detailbeschreibung	Die Variable zeigt für ASU- und XASU-Sätze den Erwerbsstatus vor der Arbeitsuche an. Ab 12/2002 erfolgte eine deutliche Reduzierung der Ausprägungen. Die Werte älterer Sätze wurden auf die aktuell gültigen Werte umgeschlüsselt, die somit für den gesamten Zeitraum gültig sind. Wird ein Datensatz innerhalb der ASU/XASU künstlich gesplittet (etwa bei einem Rechtskreiswechsel zwischen SGB III und SGB II), wird der Zugangsgrund „DS durch Splitting erzeugt“ vergeben. Dies gilt jedoch nicht für das Episodensplitting innerhalb der SIAB. Die Angabe bezieht sich daher nicht notwendigerweise auf das Beginndatum der Episode, sondern auf den Beginn des Originalzeitraums.
Qualitätshinweis	Der Anteil valider Werte (Füllgrad) steigt für XASU-Sätze zwischen 2005 und 2013 von ca. 8 % auf ca. 78 %.

### Status nach Arbeitsuche (estatnach)

Variablenlabel	Status nach Arbeitsuche
Variablenname	Estatnach
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal enthält den Status der Person nach Abgang aus der Arbeitslosigkeit. Längere Krankheitsepisoden können über dieses Merkmal identifiziert werden.</p> <p>Die Ausprägungen sind folgendermaßen unterteilt:</p> <ul style="list-style-type: none"> <li>• 1000er: Maßnahme (geförderte Beschäftigung)</li> <li>• 2000er: ungeförderte Beschäftigung</li> <li>• 3000er: in Ausbildung etc.</li> <li>• 4000er: in Selbständigkeit</li> <li>• 5000er: in Sperre</li> <li>• 6000er: in Sonstige</li> </ul>

### Profillage (profil)

Variablenlabel	Profillage
Variablenname	profil
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, MTH
Datentyp	numerisch
Hierarchie	keine
Detailbeschreibung	Das Merkmal gibt die dem Kunden beim Profiling zugewiesene Profillage wieder. Das Profiling dient der Erstellung eines Kundenprofils, d. h. einer Auflistung der arbeitsmarktrelevanten Fähigkeiten, Erfahrungen und Interessen, so dass die Verortung des Kunden auf dem Arbeitsmarkt erleichtert wird. Gegen Ende des Profilings wird zusammenfassend eine Profillage festgestellt. Hierzu wird zunächst die allgemeine Integrationsprognose des Kunden bestimmt. Zur Auswahl stehen:

	<ul style="list-style-type: none"> <li>• integrationsnah (Integration in den 1. Arbeitsmarkt innerhalb von 12 Monaten realistisch)</li> <li>• komplex (Integration in den 1. Arbeitsmarkt innerhalb von 12 Monaten nicht realistisch)</li> <li>• sonstiges</li> </ul> <p>Von der Bestimmung der Integrationsprognose ist die Zuweisung der Profillage abhängig. Den Kunden mit integrationsnaher Prognose können die Profillagen "Marktprofil", "Aktivierungsprofil" und "Förderprofil" zugewiesen werden, während die Kunden mit komplexer Prognose die Profillagen "Entwicklungsprofil", "Stabilisierungsprofil" oder "Unterstützungsprofil" erhalten. Die Auswahl der konkreten Profillage richtet sich nach dem vom Vermittler angenommenen Handlungsbedarf. Wenn eine "sonstige" Prognose vorliegt, kann bei der Profillage "Zuordnung nicht erforderlich" oder - nur für SGB-II-Kunden - "Integriert, aber hilfebedürftig" gewählt werden.</p>
Qualitätshinweis	Das Merkmal wurde 2009 eingeführt, aber bis 2006 aus anderen Merkmalen zurückgemappt. Die Qualität wird von der BA-Statistik seit 2010 als verlässlich eingestuft.

#### Art der Kündigung der letzten Tätigkeit (art\_kuend)

Variablenlabel	Art der Kündigung der letzten Tätigkeit
Variablenname	art_kuend
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, MTH
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	Dieses Merkmal beschreibt die Art der Beendigung des letzten Arbeits- oder Ausbildungsverhältnisses vor einer erneuten Arbeitsuche. Es kann daher zur Identifikation von Job-to-Job-Vermittlungen verwendet werden.

#### Arbeitszeit des Stellengesuchs (arbzeit)

Variablenlabel	Arbeitszeit des Stellengesuchs
Variablenname	arbzeit
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, MTH
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	Arbeitsuchende geben beim Vermittlungsvorgang an, welchem zeitlichen Umfang die gesuchte Stelle entsprechen sollte.

#### Restanspruch/geplante Dauer (restanspruch)

Variablenlabel	Restanspruch/geplante Dauer
Variablenname	restanspruch
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LeH, MTH

Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	<p>Je nach Datenquelle hat das Merkmal eine unterschiedliche Bedeutung.</p> <p><b>1) LeH</b>  Das Merkmal enthält den Anspruch, der nach Ende des aktuellen Leistungsbezugs übrig bleibt. Endet der Leistungsbezug vor dem Ende der maximalen Anspruchsdauer (z. B. durch Arbeitsaufnahme), bleibt ein Restanspruch mit der Dauer des nicht ausgeschöpften Anspruchs bestehen. Entsteht innerhalb von fünf Jahren ein neuer Anspruch, wird die Dauer des Restanspruchs zur neuen Anspruchsdauer addiert. Allerdings ist hierbei die Höchstdauer für das jeweilige Lebensalter die Obergrenze. Entsteht kein neuer Anspruch, so kann der Restanspruch auf Antrag innerhalb von vier Jahren für Leistungen verwendet werden.  Liegt das Enddatum des Leistungsempfangs vor dem 1.1.1998, wird der Restanspruch in Werktagen angegeben. Danach erfolgt die Angabe in Kalendertagen.  Die Angabe bezieht sich nicht auf das Enddatum der Episode, sondern auf das Ende des Originalzeitraums.</p> <p><b>2) MTH</b>  Das Merkmal enthält die geplante Dauer der Maßnahme.</p>

### Trägerart (traeger)

Variablenlabel	Trägerart
Variablenname	traeger
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal enthält die Trägerschaftsart des Trägers der Grundsicherung (LHG), des Trägers der Maßnahme (MTH) oder des Trägers der das Bewerberprofil verwaltet (ASU/XASU). Die Dimension enthält neben der SGB-III-Trägerschaftsart "Bundesagentur für Arbeit (BA)" drei Trägerschaftsarten des SGB II. Für weitere Informationen siehe Abschnitt 2.5. Die Trägerschaftsart von Grundsicherungsträgern kann im Zeitverlauf wechseln (z.B. von ARGE zu gAw oder gE zu zkT). Zu beachten ist außerdem, dass die Grenzen der Träger in Bezug auf Kreise nicht immer trennscharf sind und somit nicht mit den Grenzen der Kreise übereinstimmen müssen. Gleiches gilt für die Arbeitsagenturen. Bei ASU- und XASU-Sätzen ist die Art des Trägers enthalten, der das Bewerberangebot führt.</p>

### Beginndatum der Arbeitslosigkeit (alo\_beg)

Variablenlabel	Beginndatum der Arbeitslosigkeit
Variablenname	alo_beg
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

Herkunft	LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal gibt das Beginndatum einer ununterbrochenen Folge von Arbeitslosigkeitsperioden an und ist zu Beginn des Satzes gültig.</p> <p>Folgende Lücken führen nicht zu einer Unterbrechung der Arbeitslosigkeit:</p> <ul style="list-style-type: none"> <li>• jede Lücke kleiner oder gleich 7 Tage</li> <li>• Krankheitslücken bis zu 42 Tagen (ASU)</li> </ul> <p>Bei XASU-Sätzen sind Informationen zu Krankheit nicht enthalten, weshalb Sie bei der Berechnung nicht berücksichtigt werden können.</p>

#### Dauer der Arbeitslosigkeit (alo\_dau)

Variablenlabel	Dauer der Arbeitslosigkeit
Variablenname	alo_dau
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal gibt die Dauer (in Tagen) einer ununterbrochenen Folge von Arbeitslosigkeitsperioden an und ist zu Beginn des Satzes gültig.</p> <p>Folgende Lücken führen nicht zu einer Unterbrechung der Arbeitslosigkeit:</p> <ul style="list-style-type: none"> <li>• jede Lücke kleiner oder gleich 7 Tage</li> <li>• Krankheitslücken bis zu 42 Tagen (ASU)</li> </ul> <p>Bei der Berechnung der Dauer werden diese Lücken jedoch nicht addiert. Bei XASU-Sätzen sind Informationen zu Krankheit nicht enthalten, weshalb Sie bei der Berechnung nicht berücksichtigt werden können.</p> <p>Vor 1997 bedeutet der Wert „0“ nicht, dass keine Arbeitslosigkeit vorliegt, da hier die Quellen ASU/XASU nicht gefüllt sind.</p>

## 6.6 Ortsangaben

### Wohnort - Kreis (wo\_kreis)

Variablenlabel	Wohnort - Kreis
Variablenname	wo_kreis
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Bundesland   Kreis



Detailbeschreibung	<p>In BeH- und LeH-Sätzen liegt der Kreis des Wohnortes erst für die Jahre ab 1999 vor. Darin ist jeweils der Kreis (kreisfreie Stadt oder Landkreis) angegeben, in dem der Versicherte wohnt. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des Bundeslandes, in der ersten bis dritten Stelle den Regierungsbezirk und in der ersten bis fünften Stelle den Kreis. Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null.</p> <p>In der BeH wird der Wohnort jeweils zum Jahresende ermittelt und dann einheitlich für alle Sätze eines Jahres zugespielt. Für die Quellen LHG und XASU gilt der Wohnort für den Zeitraum der Originalbeobachtung. Bei ASU, LeH und MTH enthält das Merkmal den Wohnort zum Beginn des Originalzeitraums.</p> <p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben für alle Quellen auf den Gebietsstand vom 31.12.2014 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2014 hatten. Da sich die Grenzen der Kreise im Zeitverlauf geändert haben, würden ohne Gebietsstandsaktualisierung Fälle auftreten, bei denen die Kreis-kennziffer des Wohnortes wechselt, ohne dass der betreffende Versicherte umgezogen ist.</p>
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird als Regionalangabe lediglich das Bundesland (wo_bula) bereitgestellt.

#### Wohnort - Bundesland (wo\_bula)

Variablenlabel	Wohnort - Bundesland
Variablenname	wo_bula
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Bundesland   Kreis
Detailbeschreibung	<p>Dieses Merkmal ist eine Aggregation aus dem Merkmal Wohnort Kreis auf die 16 deutschen Bundesländer.</p> <p>In BeH- und LeH-Sätzen liegt das Bundesland des Wohnortes erst für die Jahre ab 1999 vor. Weitere Informationen zum Kreis des Wohnortes finden sich unter Wohnort - Kreis (wo_kreis).</p>

#### Wohnort - Arbeitsagentur (wo\_aa)

Variablenlabel	Wohnort - Arbeitsagentur
Variablenname	wo_aa
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Regionaldirektion   Arbeitsagentur

Detailbeschreibung	<p>Das Merkmal enthält ab 1999 die Arbeitsagentur, die am Wohnort des Beschäftigten/Kunden zuständig ist. Sie wird aus der Wohnadresse ermittelt. Für die Quellen LHG und XASU gilt der Wohnort für den Zeitraum der Originalbeobachtung. Bei LeH, ASU und MTH enthält das Merkmal den Wohnort zum Beginn der Meldung. Es ist für die BeH und LeH erst ab 1999 vorhanden.</p> <p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Arbeitsagenturangaben für alle Quellen auf den Gebietsstand vom 31.12.2014 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Arbeitsamtsbezirk nach den Grenzen, welche die Arbeitsagenturbezirke zum 31.12.2014 hatten. Da sich die Grenzen der Arbeitsagenturbezirke im Zeitverlauf geändert haben, würden ohne diese Gebietsstandsaktualisierung Fälle auftreten, bei denen der Arbeitsagenturbezirk einer Person wechselt, ohne dass diese Person umgezogen ist. Einen Problemfall stellt hinsichtlich der Gebietsstandsaktualisierung allerdings Berlin dar: Die Berliner Arbeitsagenturbezirke wurden im Laufe der Jahre immer wieder neu zugeschnitten, was auch durch die Gebietsstandsumschlüsselung nicht vollständig bereinigt werden konnte.</p> <p>Vor 1999 enthält das Merkmal die Arbeitsagentur, die den Leistungsfall zuletzt bearbeitet hat. Sie wird nicht aus der Wohnortadresse ermittelt und ist nur für die Quelle LeH gefüllt. Außerdem ist eine Gebietsstandsaktualisierung nicht möglich. Dies ist bei Berechnungen im Zeitverlauf (z. B. regionale Arbeitslosenzahlen) zu beachten.</p>
Besonderheiten	<p>Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird lediglich das Gebiet der Regionaldirektion, in dem der Wohnort des Versicherten liegt, ausgewiesen.</p>

### Wohnort - Regionaldirektion (wo\_rd)

Variablenlabel	Wohnort - Regionaldirektion
Variablenname	wo_rd
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Regionaldirektion   Arbeitsagentur
Detailbeschreibung	Dieses Merkmal ist eine Aggregation aus dem Merkmal Wohnort - Arbeitsagentur auf die Ebene der Regionaldirektionen. Weitere Informationen finden sich dort.

## 6.7 Betriebsmerkmale

### WZ73 3-Steller (w73\_3)

Variablenlabel	WZ73 3-Steller
Variablenname	w73_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch

Hierarchie	Wirtschaftsabteilung (1-Steller)   Wirtschaftsgruppe (2-Steller)   Wirtschaftsklasse (3-Steller)
Detailbeschreibung	<p>Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WS73 als 3-Steller aus und ist für die Jahre 1975 bis einschließlich 2002 gefüllt.</p> <p>WS73 steht für die "Klassifikation der Wirtschaftszweige 1973" (Bundesagentur für Arbeit, 2009). Mittels eines dreistelligen Zahlencodes werden 269 Wirtschaftsklassen unterschieden, wobei die erste Stelle dieses Codes die Wirtschaftsabteilung, insgesamt 10, definiert und die beiden ersten Stellen zusammen die jeweilige Gruppe, insgesamt 95, festlegen. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Die Zuordnung zu der entsprechenden Wirtschaftsklasse erfolgt unter Beachtung der institutionellen Ausrichtung des Betriebes.</p>

### WZ93 5-Steller (w93\_5)

Variablenlabel	WZ93 5-Steller
Variablenname	w93_5
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller)   Wirtschaftsabteilung (2-Steller)   Wirtschaftsgruppe (3-Steller)   Wirtschaftsklasse (4-Steller)   Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	<p>Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ93 als 5-Steller aus und ist für die Jahre 1999 bis inklusive 2003 gefüllt. WZ93 steht für die „Klassifikation der Wirtschaftszweige, Ausgabe 1993“ (Statistisches Bundesamt, 2002). Die WZ93 baut auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.1 („Nomenclature générale des activités économiques dans les communautés européennes“) auf, die vier Gliederungsebenen vorsieht und bezüglich der ersten beiden wiederum auf dem internationalen Standard ISIC Rev.3 („International Standard Industrial Classification of All Economic Activities“) basiert.</p> <p>Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.</p>
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w93_3) bereitgestellt.

### WZ93 3-Steller (w93\_3)

Variablenlabel	WZ93 3-Steller
Variablenname	w93_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch

Hierarchie	Wirtschaftsabschnitt (1-Steller)   Wirtschaftsabteilung (2-Steller)   Wirtschaftsgruppe (3-Steller)   Wirtschaftsklasse (4-Steller)   Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	<p>Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ93 als 3-Steller aus und ist für die Jahre 1999 bis inklusive 2003 gefüllt. WZ93 steht für die „Klassifikation der Wirtschaftszweige, Ausgabe 1993“ (Statistisches Bundesamt, 2002). Die WZ93 baut auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.1 („Nomenclature générale des activités économiques dans les communautés européennes“) auf, die vier Gliederungsebenen vorsieht und bezüglich der ersten beiden wiederum auf dem internationalen Standard ISIC Rev.3 („International Standard Industrial Classification of All Economic Activities“) basiert.</p> <p>Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.</p>

### WZ03 5-Steller (w03\_5)

Variablenlabel	WZ03 5-Steller
Variablenname	w03_5
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller)   Wirtschaftsabteilung (2-Steller)   Wirtschaftsgruppe (3-Steller)   Wirtschaftsklasse (4-Steller)   Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	<p>Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ03 als 5-Steller aus und ist für die Jahre 2003 bis 2008 gefüllt. WZ03 steht für die "Klassifikation der Wirtschaftszweige, Ausgabe 2003" (Statistisches Bundesamt, 2003). Die WZ03 baut ebenso wie die WZ93 auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.1.1 auf (siehe Variablenbeschreibung w93_3, w93_5). Die Klassifikationen der Wirtschaftszweige wurden aktualisiert, wobei die Struktur der WZ93 weitgehend beibehalten wurde.</p> <p>Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.</p>
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w03_3) bereitgestellt.

### WZ03 3-Steller (w03\_3)

Variablenlabel	WZ03 3-Steller
Variablenname	w03_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch

Hierarchie	Wirtschaftsabschnitt (1-Steller)   Wirtschaftsabteilung (2-Steller)   Wirtschaftsgruppe (3-Steller)   Wirtschaftsklasse (4-Steller)   Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	<p>Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ03 als 3-Steller aus und ist für die Jahre 2003 bis 2008 gefüllt. WZ03 steht für die "Klassifikation der Wirtschaftszweige Ausgabe 2003" (Statistisches Bundesamt, 2003). Die WZ03 baut ebenso wie die WZ93 auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.1.1 auf (siehe Variablenbeschreibung w93_3, w93_5). Die Klassifikationen der Wirtschaftszweige wurden aktualisiert, wobei die Struktur der WZ93 weitgehend beibehalten wurde.</p> <p>Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widerspiegelt werden.</p>

### WZ08 5-Steller (w08\_5)

Variablenlabel	WZ08 5-Steller
Variablenname	w08_5
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller)   Wirtschaftsabteilung (2-Steller)   Wirtschaftsgruppe (3-Steller)   Wirtschaftsklasse (4-Steller)   Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	<p>Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ08 als 5-Steller aus und ist für die Jahre ab 2008 gefüllt. WZ08 steht für die "Klassifikation der Wirtschaftszweige, Ausgabe 2008" (Statistisches Bundesamt, 2008). Die WZ08 baut auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.2 auf.</p> <p>Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widerspiegelt werden.</p>
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w08_3) bereitgestellt.

### WZ08 3-Steller (w08\_3)

Variablenlabel	WZ08 3-Steller
Variablenname	w08_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller)   Wirtschaftsabteilung (2-Steller)   Wirtschaftsgruppe (3-Steller)   Wirtschaftsklasse (4-Steller)   Wirtschaftsunterklasse (5-Steller)

Detailbeschreibung	<p>Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ08 als 3-Steller aus und ist für die Jahre ab 2008 gefüllt. WZ08 steht für die "Klassifikation der Wirtschaftszweige, Ausgabe 2008" (Statistisches Bundesamt, 2008). Die WZ08 baut auf der statistischen Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft NACE Rev.2 auf.</p> <p>Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widerspiegelt werden.</p>
--------------------	--

### W73\_3 vervollständigt durch Extrapolation/Imputation (w73\_3\_gen)

Variablenlabel	w73_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w73_3_gen
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabteilung (1-Steller)   Wirtschaftsgruppe (2-Steller)   Wirtschaftsklasse (3-Steller)
Detailbeschreibung	<p>Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WS73 als 3-Steller aus. Von 1975 bis einschließlich 2002 enthält das Merkmal die Originalwerte aus w73_3. Ab 2003 wird die Information entweder fortgeschrieben oder mit Hilfe von Umschlüsselungstabellen ersetzt. Das Merkmal bietet somit eine zeitkonsistente Wirtschaftszweigangabe auf Basis der Wirtschaftszweigklassifikation WS73. Eine ausführliche Beschreibung dazu findet sich bei Eberle et al., 2011.</p> <p>Weitere Informationen zur Klassifikation WS73 finden sich bei der Variablenbeschreibung w73_3.</p>

### Art der Vervollständigung w73\_3 (group\_w73\_3)

Variablenlabel	Art der Vervollständigung w73_3
Variablenname	group_w73_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Dieses Merkmal weist die Art der Vervollständigung für das Merkmal w73_3_gen aus. Es ist angegeben, ob der jeweilige Wert in w73_3_gen dem Originalwert von w73_3 entspricht, weiterhin fehlend ist bzw. extrapoliert oder auf Basis von Umschlüsselungstabellen imputiert wurde.</p> <p>Eine ausführliche Beschreibung zur Vorgehensweise findet sich bei Eberle et al., 2011.</p>

### W93\_3 vervollständigt durch Extrapolation/Imputation (w93\_3\_gen)

Variablenlabel	W93_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w93_3_gen
Kategorie	Betriebsmerkmale

Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller)   Wirtschaftsabteilung (2-Steller)   Wirtschaftsgruppe (3-Steller)   Wirtschaftsklasse (4-Steller)   Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	<p>Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ93 als 3-Steller aus. Von 1998 bis einschließlich 2003 enthält das Merkmal die Originalwerte aus w93_3. Vor 1998 und nach 2003 wird die Information entweder rück- bzw. fortgeschrieben oder mit Hilfe von Umschlüsselungstabellen ersetzt, sodass das Merkmal eine zeitkonsistente Wirtschaftszweigangabe auf Basis der Wirtschaftszweigklassifikation WZ93 enthält. Eine ausführliche Beschreibung dazu findet sich bei Eberle et al., 2011.</p> <p>Weitere Informationen zur Klassifikation WZ93 finden sich bei der Variablenbeschreibung w93_3.</p>

### Art der Vervollständigung w93\_3 (group\_w93\_3)

Variablenlabel	Art der Vervollständigung w93_3
Variablenname	group_w93_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Dieses Merkmal weist die Art der Vervollständigung für das Merkmal w93_3_gen aus. Es ist angegeben, ob der jeweilige Wert in w93_3_gen dem Originalwert von w93_3 entspricht, weiterhin fehlend ist bzw. extrapoliert oder auf Basis von Umschlüsselungstabellen imputiert wurde.</p> <p>Eine ausführliche Beschreibung zur Vorgehensweise findet sich bei Eberle et al., 2011.</p>

### Jahr des 1. Auftretens der Betriebsnummer (grd\_jahr)

Variablenlabel	Jahr des 1. Auftretens der Betriebsnummer
Variablenname	grd_jahr
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Bei dieser Variablen handelt es sich um das erste Auftreten der Nummer des Betriebes im Datensatz.</p>



	Wird eine Betriebsnummer in Westdeutschland erst nach 1975 bzw. in Ostdeutschland nach 1992 erstmalig nachgewiesen, könnte es sich um das Gründungsdatum des betreffenden Betriebes handeln. Allerdings könnte es sich auch um einen bereits länger existierenden Betrieb handeln, dem nach einem Wechsel des Inhabers oder einer Änderung der Rechtsform eine neue Betriebsnummer zugeordnet wurde (zur Betriebsnummernvergabe siehe Bundesagentur für Arbeit 2007, S. 9-11). Ebenfalls könnte es sein, dass der Betrieb bereits vorher existierte, aber keine sozialversicherungspflichtig bzw. ab 1999 auch keine geringfügig Beschäftigten hatte.
--	---

### Erstes Auftreten Betriebsnummer (grd\_dat)

Variablenlabel	Erstes Auftreten Betriebsnummer
Variablenname	grd_dat
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Bei dieser Variablen handelt es sich um das tagesgenaue erste Auftreten der Nummer des Betriebes in der BeH. Wird eine Betriebsnummer in Westdeutschland erst nach 1975 bzw. in Ostdeutschland nach 1992 erstmalig nachgewiesen, könnte es sich um das Gründungsdatum des betreffenden Betriebes handeln. Allerdings könnte es sich auch um einen bereits länger existierenden Betrieb handeln, dem nach einem Wechsel des Inhabers oder einer Änderung der Rechtsform eine neue Betriebsnummer zugeordnet wurde (zur Betriebsnummernvergabe siehe Bundesagentur für Arbeit, 2007, S. 9-11). Ebenfalls könnte es sein, dass der Betrieb bereits vorher existierte, aber keine sozialversicherungspflichtig bzw. ab 1999 auch keine geringfügig Beschäftigten hatte.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen tagesgenau bereitgestellt. Ansonsten wird lediglich das Jahr des ersten Auftretens der Betriebsnummer (grd_jahr) bereitgestellt.

### Jahr des letzten Auftretens der Betriebsnummer (lzt\_jahr)

Variablenlabel	Jahr des letzten Auftretens der Betriebsnummer
Variablenname	lzt_jahr
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine



Detailbeschreibung	Bei dieser Variablen handelt es sich um das letztmalige Auftreten der Nummer des Betriebes im Datensatz (vgl. Bender et al., 1996). Endet die Existenz einer Betriebsnummer im BHP bereits vor dem Jahr 2008, könnte es sich um eine Betriebsschließung handeln. Allerdings kommen als Ursache auch ein „willkürlicher Wechsel der Betriebsnummer bei Wechsel des Inhabers oder Änderung der Rechtsform“, eine „Ausgliederung von Unternehmensteilen unter einer neue Nummer“ oder andere verwaltungstechnische Änderungen in Betracht (vgl. Bender et al., 1996 oder Bundesagentur für Arbeit, 2007, S. 9-11).
--------------------	---

#### Letztes Auftreten Betriebsnummer (lzt\_dat)

Variablenlabel	Letztes Auftreten Betriebsnummer
Variablenname	lzt_dat
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Bei dieser Variablen handelt es sich um das tagesgenaue letztmalige Auftreten der Nummer des Betriebes im Datensatz (vgl. Bender et al., 1996). Endet die Existenz einer Betriebsnummer im BHP bereits vor dem Jahr 2008, könnte es sich um eine Betriebsschließung handeln. Allerdings kommen als Ursache auch ein „willkürlicher Wechsel der Betriebsnummer bei Wechsel des Inhabers oder Änderung der Rechtsform“, eine „Ausgliederung von Unternehmensteilen unter einer neue Nummer“ oder andere verwaltungstechnische Änderungen in Betracht (vgl. Bender et al., 1996 oder Bundesagentur für Arbeit, 2007, S. 9-11).
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen tagesgenau bereitgestellt. Ansonsten wird lediglich das Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr) bereitgestellt.

#### Anzahl Beschäftigte gesamt (az\_ges)

Variablenlabel	Anzahl Beschäftigte gesamt
Variablenname	az_ges
Kategorie	Generierte Betriebsmerkmale
Herkunft	BeH
Detailbeschreibung	Die Variable enthält die Anzahl aller zum 30. Juni eines Jahres zur Sozialversicherung gemeldeten Mitarbeiter eines Betriebes. Seit Einführung der neuen Meldevorschriften im Jahr 1999 sind auch die geringfügig Beschäftigten erfasst. Ebenso sind Teilzeitbeschäftigte und ruhende Beschäftigungsverhältnisse (Tagesentgelt von Null) enthalten.

#### Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az\_vz)

Variablenlabel	Anzahl Vollzeit (Normalbeschäftigte + sonstige)
Variablenname	az_vz
Kategorie	Generierte Betriebsmerkmale
Herkunft	BeH

Hierarchie	Keine
Detailbeschreibung	Die Variable enthält die Anzahl der Personen im Betrieb, die am 30. Juni eines Jahres mit den Personengruppenschlüsseln 101, 140, 143, 105, 106, 112, 118, 119, 120, 149, 201, 203, 205, 999 und YYY und als Vollzeitbeschäftigte gemeldet sind. Azubis, geringfüge Beschäftigte und Personen in Altersteilzeit werden nicht berücksichtigt.

#### Anzahl geringfügig Beschäftigte (az\_gf)

Variablenlabel	Anzahl geringfügig Beschäftigte
Variablenname	az_gf
Kategorie	Generierte Betriebsmerkmale
Herkunft	BeH
Hierarchie	Keine
Detailbeschreibung	Die Anzahl der geringfügig Beschäftigten wird anhand des Personengruppenschlüssels – Ausprägungen 109 und 209 – gebildet. Dieses Merkmal ist erst seit 1999 im Datensatz enthalten, da die geringfügig Beschäftigten erst seit diesem Jahr in das Meldeverfahren zur Sozialversicherung integriert wurden.

#### Mittelwert imputiertes Bruttotagesentgelt Vollzeitbeschäftigte (te\_imp\_mw)

Variablenlabel	Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte
Variablenname	te_imp_mw
Kategorie	Generierte Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Diese Variable enthält den Mittelwert des imputierten Bruttotagesentgeltes der Vollzeitbeschäftigten eines Betriebes. Nicht enthalten sind die Löhne der geringfügig Beschäftigten, der Azubis sowie der Personen in Altersteilzeit.</p> <p>Die Werte sind für alle Jahre in Euro angegeben.</p> <p>Gemäß den Meldevorschriften zur Sozialversicherung müssen Arbeitgeber angeben, wie hoch das beitragspflichtige Bruttoarbeitsentgelt des Beschäftigten für einen bestimmten Zeitraum war (Zeitraumentgelt). Bis Ende 1998 hatten Arbeitgeber nur den sozialversicherungspflichtigen Bruttolohn zu melden. Damit wurden nur Arbeitsentgelte erfasst, die oberhalb der Geringfügigkeits- und unterhalb der Beitragsbemessungsgrenze lagen. Seit 1999 werden im Rahmen des neuen Meldeverfahrens auch Entgelte gemeldet, die unterhalb der Geringfügigkeitsgrenze liegen. Bruttolöhne, die oberhalb der Beitragsbemessungsgrenze liegen, werden jedoch weiterhin gekappt.</p> <p>Zur Ermittlung des Bruttotageslohnes wird das Zeitraumentgelt durch die Anzahl der Kalendertage des Zeitraums geteilt. Zur Berechnung der Mittelwerte wurden diese zensierten Löhne imputiert (vgl. Abschnitt 8.2 in Schmucker et al. 2016). Diese Daten wurden dann auf Betriebsebene aggregiert.</p>

### Arbeitsort Kreis (ao\_kreis)

Variablenlabel	Arbeitsort Kreis
Variablenname	ao_kreis
Kategorie	Ortsangaben
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Bundesland   Kreis
Detailbeschreibung	<p>Angegeben ist jeweils der Kreis (kreisfreie Stadt oder Landkreis), in dem der Betrieb des Beschäftigten angesiedelt ist. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des Bundeslands, in der ersten bis dritten Stelle den Regierungsbezirk und in der ersten bis fünften Stelle den Kreis. Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null.</p> <p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben auf den Gebietsstand vom 31.12.2014 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Betriebsortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2014 hatten. Da sich die Grenzen der Kreise im Zeitverlauf geändert haben, würden ohne Gebietsstandsaktualisierung Fälle auftreten, bei denen die Kreiskennziffer des Betriebsortes wechselt, ohne dass dieser Betrieb seinen Standort verlagert hat.</p>
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird als Regionalangabe lediglich das Bundesland (ao_bula) bereitgestellt.

### Arbeitsort Bundesland (ao\_bula)

Variablenlabel	Arbeitsort Bundesland
Variablenname	ao_bula
Kategorie	Ortsangaben
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Bundesland   Kreis
Detailbeschreibung	Angegeben ist jeweils das Bundesland, in dem der Betrieb angesiedelt ist. Dieses Merkmal wird aus der Kreiskennziffer (ao_kreis) generiert. Die ersten beiden Stellen der Kreiskennziffer bezeichnen das Bundesland.

## 6.8 Verknüpfungsmerkmale

### Zustimmung zur Verknüpfung (match\_ev)

Variablenlabel	Zustimmung zur Verknüpfung
Variablenname	match_ev
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	Generiert

Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Dieses Merkmal ist nur in der Verknüpfungsdatei enthalten. Es zeigt für jede befragte Person aus der IAB-SOEP MIG an, ob sie einer Verknüpfung mit den administrativen Daten in der entsprechenden Welle schriftlich zugestimmt hat. Dieses Merkmal entspricht den Variablen <code>bdpm_p_195</code> (Welle 1, Datensatz <code>bdp_mig</code> ) und <code>bepm_p_252</code> (Welle2, Datensatz <code>bep_mig</code> ).

#### Vorliegen administrativer Personendaten (`match_admin`)

Variablenlabel	Vorliegen administrativer Personendaten
Variablenname	<code>match_admin</code>
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	Generiert
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Dieses Merkmal ist nur in der Verknüpfungsdatei enthalten. Es zeigt für jeden Befragten aus der IAB-SOEP MIG an, ob zu dieser Person auch Beobachtungen aus den administrativen Personendaten vorliegen. Bei der Verknüpfung wurden teilweise mehrere Konten zu einer Person in der Befragung gespielt. Die entsprechenden Konten sind markiert und können überprüft werden. Teilweise liegen administrative Personendaten für Personen vor ( <code>match_admin=1</code> ), die nach der Variable <code>match_ev</code> nicht zugestimmt haben ( <code>match_ev≠1</code> ). Für die Verknüpfung war jedoch das Vorliegen einer unterschriebenen Einverständniserklärung ausschlaggebend, die für diese Personen tatsächlich vorliegt.

#### Anzahl der Treffer in den administrativen Daten pro befragter Person (`match_az`)

Variablenlabel	Anzahl der Treffer in den administrativen Daten pro befragter Person
Variablenname	<code>match_az</code>
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	Generiert
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Dieses Merkmal ist nur in der Verknüpfungsdatei enthalten. Es zeigt für jede befragte Person aus der IAB-SOEP MIG an, wie viele Konten für sie in den administrativen Personendaten gefunden wurden. Bei der Verlinkung wurden teilweise mehrere <code>persnr</code> zu einer Person in der Befragung gespielt. Dies kann auch für Ankerpersonen passieren, weil eine Fortschreibung der IEB Versionen dazu führen kann, dass eine Personennummer sich in zwei Personennummern in der folgenden IEB Version aufspaltet. Für die entsprechenden Konten ist <code>match_az &gt; 1</code> . Die Konten für diese Personen sind nicht zusammengefasst und das Splitting wurde nicht angeglichen.

#### Verknüpfungsmethode (`match_typ`)

Variablenlabel	Verknüpfungsart
----------------	-----------------

Variablenname	match_typ
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	Record Linkage
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Angegeben ist, mit welchem Verfahren die Befragungsdaten den SIAB-Daten zugespielt wurden.

#### **Qualität der Verknüpfung (match\_qual)**

Variablenlabel	Verknüpfungsqualität
Variablenname	match_qual
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	Record Linkage
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Angegeben ist ein Qualitätsindikator für die Personen, die mit Hilfe probabilistischer Methoden den SIAB-Daten zugespielt wurden. Umso höher der Wert des Indikators, desto höher ist die Übereinstimmung zwischen den beiden Sätzen.

## 7 Literatur

- Bachteler, Tobias** (2008): Dokumentation Record Linkage IEB-PASS. Technische Dokumentation.
- Bender, Stefan / Hilzendege, Jürgen / Rohwer, Götz / Rudolph, Helmut** (1996): Die IAB-Beschäftigtenstichprobe 1975-1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197, Nürnberg.
- Bertat, Thomas / Dundler, Agnes / Grimm, Christopher / Kiewitt, Jochen / Schomaker, Christine / Schridde, Dr. Henning / Zemann, Dr. Christian** (2013): Neue Erhebungsinhalte 'Arbeitszeit', 'ausgeübte Tätigkeit' sowie 'Schul- und Berufsabschluss' in der Beschäftigungsstatistik. Methodenbericht, Bundesagentur für Arbeit – Statistik, URL: <http://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Methodenberichte/Beschaeftigungsstatistik/Generische-Publikationen/Methodenbericht-Neue-Erhebungsinhalte-Arbeitszeit-ausgeuebte-Taetigkeit-sowie-Schul-und-Berufsabschluss-in-der-Beschaeftigungsstatistik.pdf> (Zugriff: 18.12.2016).
- Bundesagentur für Arbeit** (Hrsg.) (2005): Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung. Ausgabe Januar 2005, Nürnberg.
- Bundesagentur für Arbeit** (Hrsg.) (2007): Handbuch für die Betriebsnummernvergabe und –pflege im Rahmen des Meldeverfahren zur Sozialversicherung. Ausgabe Dezember 2007, Nürnberg.
- Bundesagentur für Arbeit** (2009): Klassifikation der Wirtschaftszweige 1973, Nürnberg. URL: <http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Wirtschaftszweige/Klassifikation-der-Wirtschaftszweige-1973-2003/Klassifikationen-der-Wirtschaftszweige-1973-2003-Nav.html> (Zugriff: 18.12.2016)
- Bundesagentur für Arbeit** (2011): Klassifikation der Berufe 2010. Band 1: Systematischer und alphabetischer Teil mit Erläuterungen, Nürnberg. URL: <http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB2010/Printausgabe-KldB-2010-Nav.html> (Zugriff: 18.12.2016).
- Bundesanstalt für Arbeit** (1988): Klassifikation der Berufe – Systematisches und Alphabetisches Verzeichnis der Berufsbenennung, Nürnberg. URL: <http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB1975-1992/KldB1975-1992-Nav.html> (Zugriff: 18.12.2016)
- Cramer, Ulrich** (1985): Probleme der Genauigkeit der Beschäftigtenstatistik. In: Allgemeines Statistisches Archiv 69: S. 56-68.
- Deutsche BKK** (2016): Ratgeber Sozialversicherung 2016, Wolfsburg, URL: [https://www.deutschebkk.de/fileadmin/user\\_upload/microsites/arbeitgeber/medien/pdf/ratgeber-sozialversicherung-2016.pdf](https://www.deutschebkk.de/fileadmin/user_upload/microsites/arbeitgeber/medien/pdf/ratgeber-sozialversicherung-2016.pdf) (Zugriff: 18.12.2016).
- Draws, Nils** (2006): Qualitätsverbesserung der Bildungsvariable in der IAB-Beschäftigtenstichprobe 1975-2001. FDZ Methodenreport, 05/2006 (de), Nürnberg.
- Eberle, Johanna / Jacobebbinghaus, Peter / Ludsteck, Johannes/ Witter, Julia** (2011): Generation of time-consistent industry codes in the face of classification changes \* Simple heuristic based on the Establishment History Panel (BHP). FDZ Methodenreport, 05/2011 (en), Nürnberg.

- Fitzenberger, Bernd / Osikominu, Aderonke / Völter, Robert** (2006): Imputation rules to improve the education variable in the IAB employment subsample. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 126, H. 3, S. 405-436.
- Herzog, Thomas N. / Scheuren, Fritz J. / Winkler, William E.** (2007): Data Quality and Record Linkage Techniques. New York: Springer.
- Ludsteck, Johannes / Thomsen, Ulrich** (2016): Imputation of the Working Time Information for the Employment Register Data. FDZ Methodenreport 01/2016 (en), Nürnberg.
- Meinken, Holger / Koch, Iris** (2004): BA-Beschäftigtenpanel 1998-2002. Codebuch, Nürnberg.
- Paulus, Wiebke / Matthes, Britta** (2013): Klassifikation der Berufe \* Struktur, Codierung und Umsteigeschlüssel. FDZ-Methodenreport, 08/2013 (de), Nürnberg.
- Schmucker, Alexandra / Seth, Stefan / Ludsteck, Johannes / Eberle, Johanna / Ganzer, Andreas** (2016): Betriebs-Historik-Panel 1975-2014. FDZ-Datenreport, 03/2016 (de), Nürnberg.
- Schnell, Rainer / Bachteler, Tobias/ Reiher, Jörg** (2005): MTB: Ein Record-Linkage-Programm für die empirische Sozialforschung. In: ZA-Information, Nr. 56, S. 93–103.
- Statistisches Bundesamt** (2002): Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ 93), Wiesbaden. URL: <https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftsklassifikationen/Content75/KlassifikationWZ93.html> (Zugriff: 18.12.2016).
- Statistisches Bundesamt** (2003): Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003), Wiesbaden. URL: <https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftsklassifikationen/Content75/KlassifikationWZ2003.html> (Zugriff: 18.12.2016).
- Statistisches Bundesamt** (2008): Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), Wiesbaden. URL: <https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftsklassifikationen/Content75/KlassifikationWZ08.html> (Zugriff: 18.12.2016)
- Wermter, Winfried / Cramer, Ulrich** (1988): Wie hoch war der Beschäftigtenanstieg seit 1983? – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. In: Mitteilungen aus der Arbeitsmarkt – und Berufsforschung 4/88, S. 468-482.

## 8 Anhang

### 8.1 Auszählungen

Auszählungen sowie Übersichten über die einzelnen Ausprägungen und Labels der Merkmale finden sich in gesonderten Dateien unter

[http://doku.iab.de/fdz/reporte/2016/DR\\_11\\_16\\_frequencies\\_labels\\_de.zip](http://doku.iab.de/fdz/reporte/2016/DR_11_16_frequencies_labels_de.zip) oder

[http://doku.iab.de/fdz/reporte/2016/DR\\_11\\_16\\_frequencies\\_labels\\_en.zip](http://doku.iab.de/fdz/reporte/2016/DR_11_16_frequencies_labels_en.zip)



## 9 Abkürzungsverzeichnis

<b>AA</b>	Agentur für Arbeit / Arbeitsamt
<b>ALG</b>	Arbeitslosengeld
<b>ARGE</b>	Arbeitsgemeinschaft
<b>ASU</b>	Arbeitsuchendenhistorik
<b>A2LL</b>	Arbeitslosengeld II – Leistungen zum Lebensunterhalt
<b>BA</b>	Bundesagentur für Arbeit
<b>BeH</b>	Beschäftigtenhistorik
<b>BfA</b>	Bundesversicherungsanstalt für Angestellte
<b>BHP</b>	Betriebs-Historik-Panel
<b>BMAS</b>	Bundesministerium für Arbeit und Soziales
<b>coArb</b>	Computerunterstützte Arbeitsvermittlung (operatives Verfahren zur Verwaltung der Vermittlung (Altverfahren))
<b>DEÜV</b>	Verordnung über die Erfassung und Übermittlung von Daten für die Träger der Sozialversicherung – Datenerfassungs- und -übermittlungsverordnung –
<b>DEVO</b>	Zweite VO über die Erfassung von Daten für die Träger der Sozialversicherung und für die BA – Datenerfassungs-Verordnung –
<b>DÜVO</b>	Zweite VO über die Datenübermittlung auf maschinell verwertbaren Datenträgern im Bereich der Sozialversicherung und der BA – Datenübermittlungs-Verordnung –
<b>EDV</b>	Elektronische Datenverarbeitung
<b>FDZ</b>	Forschungsdatenzentrum der Bundesagentur für Arbeit am IAB
<b>FELEG</b>	Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit
<b>gAw</b>	Träger mit getrennter Aufgabenwahrnehmung
<b>gE</b>	Gemeinsame Einrichtungen
<b>gT</b>	Getrennte Trägerschaft
<b>IAB</b>	Institut für Arbeitsmarkt- und Berufsforschung
<b>IABS</b>	IAB-Beschäftigtenstichprobe
<b>IEB</b>	Integrierte Erwerbsbiografien
<b>IEBS</b>	Stichprobe der Integrierten Erwerbsbiografien
<b>ISIC</b>	International Standard Industrial Classification of All Economic Activities
<b>ITM</b>	IT- und Informationsmanagement des Instituts für Arbeitsmarkt- und Berufsforschung
<b>LIAB</b>	Linked-Employer-Employee-Daten des IAB

<b>LeH</b>	Leistungsempfängerhistorik
<b>LHG</b>	Leistungshistorik Grundsicherung
<b>LVA</b>	Landesversicherungsanstalt
<b>MTH</b>	Maßnahmeteilnahmehistorik
<b>NACE</b>	Nomenclature générale des activités économiques dans les communautés européennes
<b>Schwbg</b>	Gesetz zur Sicherung der Eingliederung Schwerbehinderter in Arbeit, Beruf und Gesellschaft – Schwerbehindertengesetz
<b>SGB</b>	Sozialgesetzbuch
<b>SIAB</b>	Stichprobe der Integrierten Arbeitsmarktbiografien
<b>VerBIS</b>	Vermittlungs- und Beratungsinformationssystem
<b>XASU</b>	Arbeitsuchendenhistorik aus XSozial-BA-SGB II
<b>zkT</b>	Zugelassene kommunale Träger

## Impressum

### FDZ-Datenreport 11/2016

#### Herausgeber

Forschungsdatenzentrum (FDZ)  
der Bundesagentur für Arbeit  
im Institut für Arbeitsmarkt- und Berufsforschung  
Regensburger Str. 104  
90478 Nürnberg

#### Redaktion

Dana Müller, Heiner Frank

#### Technische Herstellung

Heiner Frank

#### Rechte

Nachdruck - auch auszugsweise - nur mit  
Genehmigung des FDZ gestattet

#### Bezugsmöglichkeit

[http://doku.iab.de/fdz/reporte/2016/DR\\_11-16.pdf](http://doku.iab.de/fdz/reporte/2016/DR_11-16.pdf)

#### Internet

<http://fdz.iab.de/>

#### Rückfragen zum Inhalt an:

Parvati Trübswetter  
Institut für Arbeitsmarkt- und  
Berufsforschung (IAB)  
Forschungsbereich B1  
Regensburger Str. 104  
90478 Nürnberg  
Telefon: +49 (0) 911/179-3230  
E-Mail: [Parvati.Truebswetter@iab.de](mailto:Parvati.Truebswetter@iab.de)