



FORSCHUNGSDATENZENTRUM  
der Bundesagentur für Arbeit im Institut für  
Arbeitsmarkt- und Berufsforschung

# FDZ-DATENREPORT

Dokumentation zu Arbeitsmarktdaten

---

**10|2020 DE** NEPS-SC5-Erhebungsdaten verknüpft mit  
administrativen Daten des IAB (NEPS-SC5-ADIAB)

Nadine Bachbauer & Clara Wolf



**Bundesagentur für Arbeit**

# NEPS-SC5-Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC5- ADIAB)

Nadine Bachbauer (IAB, LIfBi) & Clara Wolf (IAB, LIfBi)

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>7</b>
1.1	Einleitung	7
1.2	Datennutzung	8
1.2.1	Datenzugang	8
1.2.2	Datenhaltung	9
1.2.3	Sensible Merkmale und Zusatzvariablen	10
1.2.4	Dateinamen der administrativen Daten und der Verknüpfungsdatei	11
1.2.5	Dateinamen der NEPS-Startkohorte 5	12
1.3	Steckbrief	14
1.4	Merkmalsliste der administrativen Daten	18
	Grund Abgabe Beschäftigungsmeldung / Grund Ende Leistungsbezug (grund)	20
1.5	Mengengerüst	22
<b>2</b>	<b>Datenquellen</b>	<b>22</b>
2.1	NEPS-SC5-Erhebungsdaten (NEPS-SC5)	22
2.2	Administrative Daten aus den Integrierten Erwerbsbiografien (IEB)	23
2.3	Beschäftigtenhistorik (BeH)	24
2.4	Leistungsempfängerhistorik (LeH)	25
2.5	Leistungshistorik Grundsicherung (LHG)	25
2.6	Arbeitsuchendenhistorik (ASU und XASU)	26
2.7	Maßnahmeteilnahmehistoriken (MTH und XMTH)	26
2.8	Veränderungen zur SIAB 7514	26
2.8.1	Korrektur der Fehlbuchungen in der Gemeinde Lahn	26
2.8.2	Aufnahme der Maßnahmeteilnahmehistorik aus XSozial_BA-SGB II (XMTH)	27
2.8.3	Merkmalskanon	27
<b>3</b>	<b>Datenaufbereitung und Stichprobenziehung</b>	<b>28</b>
3.1	Bereinigungen und Aufbereitungen	28
3.1.1	NEPS-SC5-Erhebungsdaten	28
3.1.2	Administrative Daten aus den Integrierten Erwerbsbiographien (IEB)	28
3.1.3	Beschäftigtenhistorik (BeH)	28
3.1.4	Leistungsempfängerhistorik (LeH)	29
3.1.5	Leistungshistorik Grundsicherung (LHG)	29
3.1.6	Arbeitsuchendenhistorik (ASU und XASU)	30
3.1.7	Maßnahmeteilnahmehistoriken (MTH)	30
3.1.8	Maßnahmeteilnahmehistoriken aus XSozial (XMTH)	30
3.1.9	SGB-II-Anonymisierung	31
3.2	Episodensplitting	31

3.3	Fehlende Werte .....	33
<b>4</b>	<b>Datenverknüpfung.....</b>	<b>33</b>
4.1	Informierte Zustimmung der Befragten zur Datenverknüpfung.....	33
4.2	Herkunft der Verknüpfungsmerkmale .....	33
4.3	Bereinigung und Standardisierung personenbezogener Merkmale.....	34
4.4	Exakter Abgleich der Verknüpfungsmerkmale .....	35
4.5	Fehlertoleranter Abgleich der Verknüpfungsmerkmale.....	35
4.6	Ergebnis der Datenverknüpfung .....	37
<b>5</b>	<b>Datenqualität und Probleme .....</b>	<b>38</b>
5.1	NEPS-SC5-Befragungsdaten (NEPS-SC5) .....	38
5.2	Integrierte Erwerbsbiografien (IEB).....	38
5.3	Beschäftigtenhistorik (BeH).....	39
5.4	Leistungsempfängerhistorik (LeH).....	41
5.5	Leistungshistorik Grundsicherung (LHG) .....	41
5.6	Arbeitsuchendenhistorik (ASU/XASU) .....	42
5.6.1	ASU.....	42
5.6.2	XASU.....	43
5.7	Maßnahmeteilnehmergehistoriken (MTH/XMTH) .....	44
5.7.1	MTH.....	44
5.7.2	XMTH.....	44
<b>6</b>	<b>Variablenbeschreibung.....</b>	<b>45</b>
6.1	Identifikatoren .....	45
6.1.1	Personen-ID (ID_t).....	45
6.1.2	Betriebs-ID (betnr).....	45
6.2	Generierte technische Merkmale .....	46
6.2.1	Satzzähler pro Konto (spell) .....	46
6.2.2	Quelle des Satzes (quelle).....	46
6.2.3	Jahr (jahr) .....	47
6.3	Gültigkeitszeitraum .....	47
6.3.1	Beginndatum Originalbeobachtung (begorig).....	47
6.3.2	Enddatum Originalbeobachtung (endorig).....	47
6.3.3	Beginndatum der Episode (begepi).....	48
6.3.4	Enddatum der Episode (endepi).....	48
6.4	Informationen zur Person.....	48
6.4.1	Geschlecht (frau) .....	48
6.4.2	Geburtsjahr (gebjahr).....	49
6.4.3	Geburtsmonat (gebmon) .....	49
6.4.4	Staatsangehörigkeit (nation).....	49

6.4.5	Staatsangehörigkeit vergrößert (nation_gr).....	50
6.4.6	Familienstand (famst).....	50
6.4.7	Kinderzahl (kind).....	50
6.4.8	Ausbildung (ausbildung).....	51
6.4.9	Ausbildung (imputiert) (ausbildung_imp).....	53
6.4.10	Schulausbildung (schule).....	54
6.5	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche .....	55
6.5.1	Tagesentgelt / täglicher Leistungssatz (tentgelt) .....	55
6.5.2	Beruf – ausgeübte / letzte Tätigkeit (KldB 1988) (beruf).....	56
6.5.3	Berufsgruppe – ausgeübte / letzte Tätigkeit (KldB 2010) (beruf2010_3) .....	56
6.5.4	Berufsuntergruppe – ausgeübte / letzte Tätigkeit (KldB 2010) (beruf2010_4) .....	57
6.5.5	Anforderungsniveau – ausgeübte / letzte Tätigkeit (KldB 2010) (niveau).....	58
6.5.6	Teilzeit (teilzeit).....	59
6.5.7	Stellung im Beruf und Arbeitszeit (stib).....	59
6.5.8	Erwerbsstatus (erwstat).....	60
6.5.9	Gleitzone (gleitz) .....	61
6.5.10	Leiharbeit (leih) .....	62
6.5.11	Befristung (befrist) .....	62
6.5.12	Grund Abgabe Beschäftigungsmeldung / Grund Ende Leistungsbezug (grund) ..	62
6.5.13	Erwerbsstatus vor Arbeitsuche (estatvor) .....	64
6.5.14	Status nach Arbeitsuche (estatnach) .....	64
6.5.15	Profillage (profil) .....	64
6.5.16	Art der Kündigung der letzten Tätigkeit (art_kuend).....	65
6.5.17	Arbeitszeit des Stellengesuchs (arbzeit) .....	65
6.5.18	Restanspruch / geplante Dauer (restanspruch) .....	65
6.5.19	Trägerart (traeger) .....	66
6.5.20	Beginndatum der Arbeitslosigkeit (alo_beg) .....	66
6.5.21	Dauer der Arbeitslosigkeit (alo_dau).....	67
6.6	Ortsangaben.....	67
6.6.1	Wohnort – Kreis (wo_kreis).....	67
6.6.2	Wohnort – Bundesland (wo_bula).....	68
6.6.3	Wohnort – Arbeitsagentur (wo_aa) .....	68
6.6.4	Wohnort – Regionaldirektion (wo_rd).....	69
6.7	Betriebsmerkmale .....	69
6.7.1	WZ73 3-Steller (w73_3) .....	69
6.7.2	WZ93 5-Steller (w93_5) .....	70
6.7.3	WZ93 3-Steller (w93_3) .....	70
6.7.4	WZ03 5-Steller (w03_5) .....	71
6.7.5	WZ03 3-Steller (w03_3) .....	71
6.7.6	WZ08 5-Steller (w08_5) .....	72
6.7.7	WZ08 3-Steller (w08_3) .....	72
6.7.8	w73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen).....	72
6.7.9	Art der Vervollständigung w73_3 (group_w73_3).....	73

6.7.10	w93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen).....	73
6.7.11	Art der Vervollständigung w93_3 (group_w93_3).....	73
6.7.12	w08_3 vervollständigt durch Extrapolation/Imputation (w08_3_gen).....	74
6.7.13	Art der Vervollständigung w08_3 (group_w08_3).....	74
6.7.14	Jahr des 1. Auftretens der Betriebsnummer (grd_jahr).....	74
6.7.15	Erstes Auftreten Betriebsnummer (grd_dat).....	75
6.7.16	Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr).....	75
6.7.17	Letztes Auftreten Betriebsnummer (lzt_dat) .....	76
6.7.18	Anzahl Beschäftigte gesamt (az_ges) .....	76
6.7.19	Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz) .....	76
6.7.20	Anzahl geringfügig Beschäftigte (az_gf) .....	77
6.7.21	Mittelwert imputiertes Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw) ...	77
6.7.22	Arbeitsort Kreis (ao_kreis) .....	77
6.7.23	Arbeitsort Bundesland (ao_bula) .....	78
6.8	Verknüpfungsmerkmale .....	78
6.8.1	Vorliegen administrativer Personendaten (match_admin) .....	78
6.8.2	Verknüpfungsmethode (match_typ) .....	78
6.8.3	Qualität der Verknüpfung (quality).....	79
<b>7</b>	<b>Literatur .....</b>	<b>80</b>
<b>8</b>	<b>Anhang .....</b>	<b>82</b>
8.1	Auszählungen.....	82
8.2	Abkürzungsverzeichnis.....	82

# Zusammenfassung

Dieser Datenreport beschreibt die verknüpften Erhebungsdaten des Nationalen Bildungspanels der Startkohorte 5 „Hochschulstudium und Übergang in den Beruf“ (NEPS-SC5) mit administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB).

## Abstract

This data report describes the linked survey data of Starting Cohort 5 „From Higher Education to the Labor Market“ of the National Educational Panel Study (NEPS-SC5) with administrative data of the Institute for Employment Research (IAB).

## Keywords

German administrative micro data, labour market data, data manual, National Educational Panel Study, survey data

## 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 das Datenprodukt „NEPS-SC5-Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC5-ADIAB 7518)“, das gemeinsam vom Leibniz-Institut für Bildungsverläufe e.V. (LIfBi) und dem Institut für Arbeitsmarkt- und Berufsforschung (IAB) angeboten wird. Die verknüpften Daten bestehen aus in Interviews erhobenen Informationen von Teilnehmerinnen und Teilnehmern der sogenannten Studierendenkohorte (SC5) des Nationalen Bildungspanels (NEPS) und deren Personendaten aus den administrativen Daten des IAB, sofern die Befragten ihre Zustimmung zur Verknüpfung der Daten erteilt haben und in den Daten des IAB identifiziert werden konnten.

Die Daten des NEPS wurden von 2008 bis 2013 als Teil des Rahmenprogramms zur Förderung der empirischen Bildungsforschung erhoben, welches vom Bundesministerium für Bildung und Forschung (BMBF) finanziert wurde. Seit 2014 wird das NEPS vom Leibniz-Institut für Bildungsverläufe e.V. (LIfBi) an der Otto-Friedrich-Universität Bamberg in enger Kooperation mit einem deutschlandweiten Exzellenznetzwerk von Partnerinstituten weitergeführt. Weitere Hintergrundinformationen dazu finden sich auf der Website des NEPS<sup>1</sup>.

Inhaltlich zielt die Startkohorte 5 des Nationalen Bildungspanels darauf ab, Studierende auf ihrem Weg durch das Studium bis in den Beruf hinein zu begleiten. Ein besonderer Schwerpunkt liegt dabei auf der Erhebung fachlicher und außerfachlicher Kompetenzen von Studierenden sowie der Kompetenzentwicklung innerhalb des Studienverlaufs. Die Studie ermöglicht es damit den Einfluss von Kompetenzen auf den Studienerfolg und den beruflichen Erfolg von Studierenden zu ermitteln. Darüber hinaus erlauben die Daten Bildungsentscheidungen und deren Bestimmungsfaktoren ebenso wie deren Konsequenzen für den weiteren Bildungs- und Berufsweg der Studierenden zu untersuchen. Der Übergang in den Arbeitsmarkt, der berufliche Erfolg und die Erträge von Hochschulbildung bilden ein weiteres Schwerpunktthema der Befragung.

Die Stichprobe der NEPS-SC5 beinhaltet Studienanfängerinnen und -anfänger, die im Wintersemester 2010/2011 zum ersten Mal an einer öffentlichen oder staatlich anerkannten Hochschule in Deutschland für einen Studiengang, der einen Bachelorabschluss, ein Staatsexamen in Medizin, Jura oder Lehramt, einen Diplom- oder Masterabschluss in katholischer oder evangelischer Theologie oder einen künstlerischen Abschluss verleiht, eingeschrieben waren. Studierende an Hochschulen von Bundesministerien und Bundesländern für Mitarbeiter der öffentlichen Verwaltung und des öffentlichen Dienstes wurden dabei nicht berücksichtigt. Das Sample basiert auf einer geschichteten Klumpenstichprobe, wobei alle Studierenden eines Studienbereichs an einer bestimmten Hochschule einen Klumpen darstellen. Dabei wurde ein Oversample von Lehramtsstudierenden und Studierenden an Privathochschulen in das Studierendenpanel implementiert. Zusätzlich wurde eine Vollerhebung von nicht-traditionellen Studienanfängerinnen und -anängern angestrebt.

---

<sup>1</sup> Siehe <https://www.neps-data.de>.

Um Zugang zu den Studierenden zu erhalten, wurden alle Hochschulen der auf diese Weise ausgewählten Klumpen durch das Erhebungsinstitut kontaktiert. Die darauffolgende Rekrutierung erfolgte auf zwei Wegen: Zum einen wurden die Studierenden durch die Hochschulverwaltung schriftlich informiert, zum anderen besuchten Feldarbeiter zentrale Erstsemesterveranstaltungen der gezogenen Studienbereiche und baten die Studierenden um ihre Teilnahme.

Für das verknüpfte Datenprodukt steht ausschließlich die Download-Version der NEPS-Daten zur Verfügung. Die Remote- bzw. Onsite-Version, deren Daten weniger stark anonymisiert sind, können für NEPS-SC5-ADIAB nicht bereitgestellt werden.

Die administrativen Daten stammen aus den Integrierten Erwerbsbiografien (IEB) des IAB. Sie enthalten im vorliegenden Fall Daten zu allen Personen, die im Rahmen der NEPS Startkohorte 5 befragt wurden, einer Verknüpfung zugestimmt haben und verknüpft werden konnten, sofern sie 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 1997),
- Teilnahme an einer arbeitsmarktpolitischen Maßnahme der BA (erfasst ab 2000).

Diese aus unterschiedlichen Datenquellen stammenden Informationen werden in den IEB zusammengeführt, tagesgenau abgebildet und 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 Teilnahmen an arbeitsmarktpolitischen Maßnahmen in der Maßnahmeteilnehmergehistorik (MTH und XMTH) erfasst werden.

Neben diesen personenbezogenen Daten können auch Angaben zu den beschäftigenden Betrieben aus dem Betriebs-Historik-Panel (BHP) des IAB (Schmucker et al., 2018) zugespielt werden.

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 der administrativen Daten. 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

### 1.2.1 Datenzugang

Die NEPS-SC5-ADIAB-Daten sind schwach anonymisiert und können ausschließlich im Rahmen von Gastaufenthalten an einem der Standorte des Forschungsdatenzentrums der Bundesagentur

für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (IAB-FDZ) und anschließender Datenfernverarbeitung ausgewertet werden.

Für die Datennutzung ist in einem ersten Schritt ein Antrag beim IAB-FDZ einzureichen. Über die Genehmigung des Forschungsprojekts entscheidet das FDZ im Auftrag und ggf. unter Abstimmung mit dem Bundesministerium für Arbeit und Soziales (BMAS). Nach erfolgter Genehmigung muss ein Datennutzungsvertrag für die NEPS-Daten mit dem Forschungsdatenzentrum des LfBi (LfBi-FDZ) geschlossen werden.<sup>2</sup> Sobald der Datennutzungsvertrag für die NEPS-Daten dem IAB vorliegt, wird ein Nutzungsvertrag zwischen der Institution der Forscherin bzw. des Forschers und dem IAB geschlossen. Details zur Beantragung des Datensatzes und den Datenverarbeitungsmöglichkeiten können der Website des IAB-FDZ entnommen werden. Der Datenzugriff beginnt somit erst, wenn die antragstellende Person nachweislich Verträge mit beiden Forschungsdatenzentren abgeschlossen hat.

Zur Beantwortung von Fragen zu diesem Datenprodukt können sich Nutzerinnen und Nutzer an die Mitarbeiterinnen und Mitarbeiter des IAB-FDZ und des LfBi-FDZ wenden. Die Arbeitsteilung der beiden Einrichtungen sieht vor, dass sich die Nutzerinnen und Nutzer mit Fragen zu den NEPS-Daten ausschließlich an das LfBi-FDZ wenden, mit Fragen zu den administrativen Daten, zu technischen Aspekten der Datenverknüpfung oder des Zugriffs auf die Daten ausschließlich an das IAB-FDZ. Bei allgemeinen Fragen zum NEPS kann neben dem LfBi-FDZ auch das NEPS-Forum<sup>3</sup> konsultiert werden.

### 1.2.2 Datenhaltung

Die Erhebungsdaten des NEPS und die administrativen Daten des IAB sind in getrennten Dateien abgelegt, was die Struktur der Daten verdeutlicht und Speicherplatz in der Datenhaltung spart. Die Personendaten aus beiden Datenquellen lassen sich über die NEPS-Personen-ID (ID\_t) zusammenführen, die auch in den administrativen Personendaten enthalten ist.

Ein weiteres Element von NEPS-SC5-ADIAB ist die Verknüpfungsdatei, welche die Personnummer sowie technische Merkmale mit Informationen darüber beinhaltet, ob bzw. wie die Verknüpfung mit den administrativen Daten gelungen ist. Die Datei umfasst auch die Befragten der NEPS Startkohorte 5, zu denen keine administrativen Daten gefunden werden konnten. Somit können Ausfallanalysen auf Basis aller Teilnehmerinnen und Teilnehmer der NEPS-SC5 durchgeführt werden.

Die administrativen Daten, die analog zu den NEPS-Daten, sowohl deutsche als auch englische Labels enthalten<sup>4</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. Diese Personendaten umfassen tagesgenaue Angaben für den Zeitraum 1975 bis 2017.

---

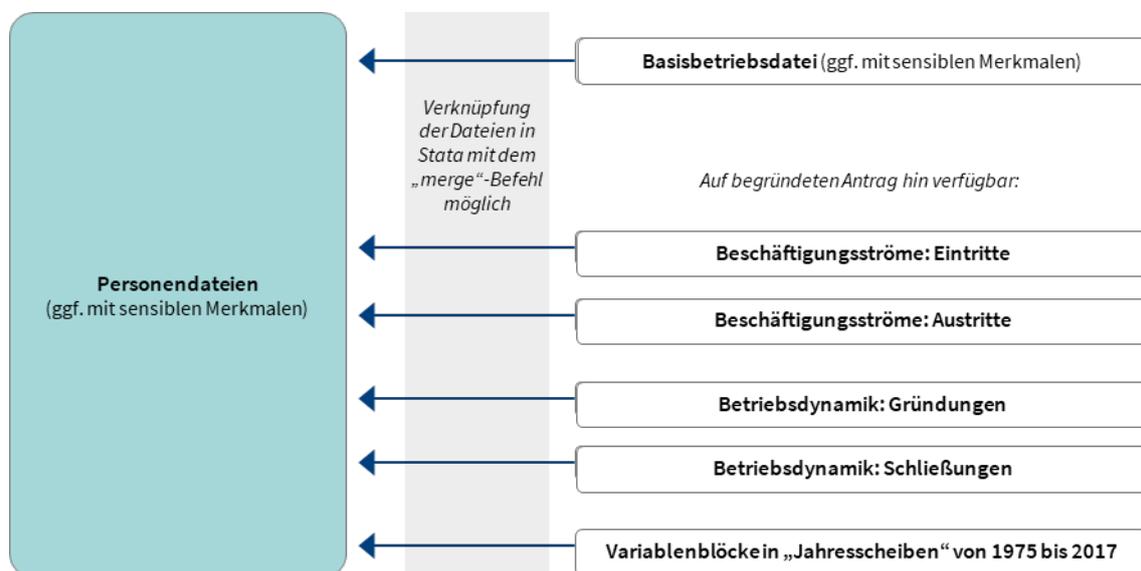
<sup>2</sup> Siehe <https://www.neps-data.de/Datenzentrum/Datenzugang/Datennutzungsverträge>.

<sup>3</sup> Siehe <https://forum.lifbi.de>.

<sup>4</sup> Mit den Stata-Befehlen `label language en` bzw. `label language de` kann auf die englisch- bzw. deutschsprachigen Labels umgestellt werden.

Ein zweites Modul, die Basisbetriebsdatei, enthält die Betriebsnummer, das Jahr, Angaben zu Arbeitsort und Wirtschaftszweigen sowie weitere Betriebsinformationen zum Stichtag 30. Juni der Jahre 1975 bis 2017. Somit sind administrative Betriebs- und Personenmerkmale in getrennten Dateien abgelegt (vgl. Abbildung 1). Zusätzlich sind in der Basisbetriebsdatei generierte Variablen für über die Zeit konsistente Wirtschaftszweigklassifikationen enthalten (s. Eberle et al., 2011). Zudem können weitere Betriebsvariablen und Erweiterungsmodule des Betriebs-Historik-Panel (BHP) genutzt werden, die von Schmucker et al. (2018) beschrieben werden. Die Verknüpfung der administrativen Personen- und Betriebsdateien 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.

Abbildung 1 Datenhaltung der administrativen Daten



### 1.2.3 Sensible Merkmale und Zusatzvariablen

Bestimmte Merkmale in den administrativen Daten, 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:

Administrative Personendatei:

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

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)

Es ist zu beachten, dass die sensiblen Merkmale in der Regel in vergrößerter Version bereits in den Daten enthalten sind. Die sensiblen Merkmale werden nur dann detailliert weitergegeben, wenn die Information im vergrößerten Zustand zur Erreichung des Forschungsziels nicht ausreichend ist.

Auf begründeten Antrag können zudem aus dem Betriebs-Historik-Panel (BHP) weitere Betriebsmerkmale in Variablenblöcken bereitgestellt werden (vgl. Abbildung 1). Nähere Informationen zu den ggf. zusätzlich zu beantragenden Variablenblöcken aus dem BHP finden sich unter [http://fdz.iab.de/de/FDZ\\_Overview\\_of\\_Data/working\\_tools.aspx](http://fdz.iab.de/de/FDZ_Overview_of_Data/working_tools.aspx).

Auch die NEPS-Daten liegen in unterschiedlichen Anonymisierungsgraden vor. Bei den für NEPS-SC5-ADIAB verfügbaren Daten handelt es sich um die Download-Version, welche den höchsten Anonymisierungsgrad aufweist. Diese Version enthält einige Variablen, jeweils im Namen mit den Suffixen „\_R“ bzw. „\_O“ gekennzeichnet, deren Ausprägungen durch Aggregation oder andere Maßnahmen anonymisiert wurden. Die vollständigen Informationen dieser Variablen sind nur in der Remote- bzw. Onsite-Version verfügbar. Aus Datenschutzgründen können weder die Remote- noch die Onsite-Version für NEPS-SC5-ADIAB bereitgestellt werden, auch nicht durch einen gesonderten Antrag. Es besteht somit keine Möglichkeit, Variablen aus der NEPS-Erhebung zu nutzen, die nicht in der Download-Version enthalten sind. Dies betrifft vornehmlich Regionaldaten einschließlich der Informationen zum Bundesland, Länderkennzeichnungen wie Muttersprache oder Herkunftsland sowie Informationen zum Hochschulkontext.

## 1.2.4 Dateinamen der administrativen Daten und der Verknüpfungsdatei

### Personendatei

NEPS-SC5-ADIAB\_7518\_v1.dta

### Basisbetriebsdatei

NEPS-SC5-ADIAB\_7518\_v1\_bhp\_basis\_v1.dta

### Verknüpfungsdatei

NEPS-SC5-ADIAB\_7518\_v1\_key.dta

### Zusatzmodule BHP

Bestandsdaten mit Variablenblöcken

NEPS-SC5-ADIAB\_7518\_v1\_bhp\_v1\_####.dta, #### = 1975 - 2017

Beschäftigtenströme

NEPS-SC5-ADIAB\_7518\_v1\_bhp\_inflow\_v1.dta

NEPS-SC5-ADIAB\_7518\_v1\_bhp\_outflow\_v1.dta

Betriebsdynamik

NEPS-SC5-ADIAB\_7518\_v1\_bhp\_entry\_v1.dta

NEPS-SC5-ADIAB\_7518\_v1\_bhp\_exit\_v1.dta

## 1.2.5 Dateinamen der NEPS-Startkohorte 5

Tabelle 1 Datenstruktur NEPS-SC5

Datensatz	Dateiname	Inhalt und Besonderheiten
<b>Basisdaten</b>	Basics.dta	Informationen zu aktuellen Basismerkmalen der Befragten (z.B. Soziodemographie, aktuelle Erwerbstätigkeits- und Haushaltscharakteristika); Querschnittsdaten im Wide-Format
<b>Profildaten</b>	CohortProfile.dta	Informationen zur Teilnahme, zum Interviewdatum, zur Verfügbarkeit von Befragungs- und Kompetenzdaten; Ausgangsdatei für das Zusammenspielen von Informationen (Identifikatoren); Längsschnittdaten im Long-Format
<b>Methodendaten zu Befragungsdaten</b>	MethodsCATI.dta	Informationen zum Feldverlauf und zur Datenerhebung der CATI-Befragungen (Rücklauf, Kontaktierung, Dauer, Interviewer, Incentives); Längsschnittdaten im Long-Format
	MethodsCAWI.dta	Informationen zum Feldverlauf und zur Datenerhebung der CAWI-Befragungen (Rücklauf, Kontaktierung, Dauer, Incentives); Längsschnittdaten im Long-Format
<b>Gewichtungsdaten</b>	Weights.dta	Informationen zu Stichprobencharakteristika und Stratifizierung inklusive Gewichtungsvariablen; Längsschnittdaten im Wide-Format
<b>Befragungsdaten</b>	pTargetCATI.dta	Informationen aus den CATI-Befragungen der Zielperson; Längsschnittdaten im Long-Format
	pTargetCAWI.dta	Informationen aus den CAWI-Befragungen der Zielperson; Längsschnittdaten im Long-Format
<b>Kompetenz- und CAPI-Befragungsdaten</b>	xEcoCAPI.dta	Informationen zu den in Welle 7 durchgeführten wirtschaftswissenschaftlichen Kompetenztestungen sowie zum Studium und der Datenerhebung der CAPI-Befragungen (Rücklauf, Kontaktierung, Dauer, Interviewer, Incentives) von Studierenden eines wirtschaftswissenschaftlichen Fachs; Querschnittsdaten im Wide-Format
<b>Kompetenzdaten</b>	xTargetCompetencies.dta	Informationen zu den in Welle 1, 5 und 12 durchgeführten Kompetenztestungen; Längsschnittdaten im Wide-Format
	xPlausibleValues.dta	Plausible Values für die in xTargetCompetencies enthaltenen Kompetenzdaten; Längsschnittdaten im Wide-Format
<b>Methodendaten zu Kompetenzdaten</b>	MethodsCompetencies.dta	Informationen zur methodischen Durchführung der Kompetenztestungen; Längsschnittdaten im Long-Format
<b>Lebensverlaufs-Spelledaten</b>	Biography.dta	Integrierte und geglättete Lebensverlaufsdaten aus verschiedenen Lebensbereichen; Zusammenfassung aller vollständigen, harmonisierten und rechtszentrierten Episoden-Spells
<b>Schul- und Ausbildungsdaten</b>	Education.dta	Informationen zu den Übergängen in der (Aus-)Bildungskarriere der Zielperson (z.B. CASMIN, ISCED-

Datensatz	Dateiname	Inhalt und Besonderheiten
		97); aus den Spell-Datensätzen zur Schul- und Ausbildung generierter Datensatz
<b>Schulbildungs-Spelldaten</b>	spSchool.dta	Informationen zur Allgemeinbildungshistorie, von der Einschulung bis zum Zeitpunkt des Abschlusses; Regionaldaten (Land, Ost-/Westdeutschland)
<b>Berufsvorbereitung-Spelldaten</b>	spVocPrep.dta	Informationen zu Episoden der Berufsvorbereitung nach der allgemeinen Schulbildung
<b>Ausbildung-Spelldaten</b>	spVocTrain.dta	Informationen zu allen weiterführenden Ausbildungen, die die Zielperson jemals absolviert hat (berufliche Ausbildung, Studium); Regionaldaten (Land, Ost-/Westdeutschland)
<b>Erwerbstätigkeit-Spelldaten</b>	spEmp.dta	Informationen zu Episoden regulärer Erwerbstätigkeiten inklusive Referendariate und studentischer Erwerbstätigkeiten; Beginn einer neuen Episode bei Arbeitgeber-, Tätigkeitswechsel oder Unterbrechung der Erwerbstätigkeit; Regionaldaten (Land, Ost-/Westdeutschland)
<b>Erwerbslosigkeit-Spelldaten</b>	spUnemp.dta	Informationen zu Episoden von Erwerbslosigkeit unabhängig von der Registrierung als arbeitslos
<b>Wehr- und Zivildienst-Spelldaten</b>	spMilitary.dta	Informationen zu Episoden von Wehr- und Zivildienst als auch Zeiträume, die für Freiwilligenarbeit im sozialen Bereich oder dem Umweltsektor genutzt wurden
<b>Praktikum-Spelldaten</b>	spInternship.dta	Informationen zu Episoden von Pflichtpraktika und freiwilligen Praktika
<b>Partner-Spelldaten</b>	spPartner.dta	Informationen zur Historik von Partnerschaften der Zielperson (Zusammenleben, Heirat, Trennung, Scheidung); Merkmale des Partners (z.B. Geburtsdatum, Ausbildung, Beruf)
<b>Geschwister-Spelldaten</b>	spSibling.dta	Soziodemographische Informationen zu allen in Welle 1 berichteten Geschwistern der befragten Zielperson
<b>Kinder-Spelldaten</b>	spChild.dta	Informationen zu biologischen, Pflege- und Adoptivkindern der Zielperson sowie zu allen Kindern, die mit der Zielperson zusammenleben oder gelebt haben
<b>Elternzeiten-Spelldaten</b>	spParLeave.dta	Information über Episoden zu Elternzeiten der Zielperson bezüglich der Kinder aus spChild (ausgenommen verstorbene Kinder)
<b>Zusammenleben mit Kindern-Spelldaten</b>	spChildCohab.dta	Informationen zur Historik des Zusammenlebens der Zielperson mit Kindern
<b>Lücken-Spelldaten</b>	spGap.dta	Informationen zu Lücken in der Erfassung des individuellen Lebensverlaufs; identifiziert durch ein Prüfmodul
<b>(Berufliche) Kurse-Spelldaten</b>	spCourses.dta	Informationen zu den innerhalb der letzten 12 Monate besuchten Kursen und Weiterbildungen während einer Erwerbstätigkeits-, Erwerbslosigkeits-, Elternzeit-, Wehr-/Zivildienst- oder Lückenepisode
<b>Weitere Kurse-Spelldaten</b>	spFurtherEdu1.dta	Informationen zu Kursen, die neben den in spCourses und spVocTrain berichteten Kursen innerhalb der letzten 12 Monate besucht wurden; sowohl berufliche Weiterbildungen als auch Kurse mit privaten Zielen (z.B. Yoga, Kochkurse)
	spFurtherEdu2.dta	Zusätzlich erhobene Informationen zu zwei zufällig ausgewählten Kursen aus spCourses und spFurtherEdu1 (z.B. Kosten, Motivation, Zertifikate)

Datensatz	Dateiname	Inhalt und Besonderheiten
Schulabschlüsse im Ausland-Spelldaten	spSchoolExtExam.dta	Informationen zu Schulabschlusszertifikaten, die außerhalb des regulären deutschen Schulsystems erworben wurden
Ausbildungsabschlüsse im Ausland-Spelldaten	spVocExtExam.dta	Informationen zu Ausbildungsabschlusszertifikaten, die außerhalb des regulären deutschen Berufsausbildungssystems erworben wurden
Editionsdaten	EditionBackups.dta	Informationen zu vorgenommenen Editionen in den Datensätzen

Welche Variablen in den jeweiligen Datensätzen als eindeutige Identifikatoren zum Zusammenspielen von Informationen fungieren, kann der Merging-Matrix auf der NEPS-Website entnommen werden. Weitere Informationen zu den einzelnen Datensätzen finden sich im Data Manual der SC5, welches ebenfalls von der Webseite bezogen werden kann.

### 1.3 Steckbrief

Tabelle 2 Steckbrief NEPS-SC5-ADIAB

Themen / Merkmalsgruppen	<p><u>Erhebungsdaten</u> Längsschnittdaten zu Soziodemographie; Bildungs- und Erwerbsbiographie, Erwerbseinstieg; Kompetenzentwicklungen, Bildungsprozessen, Bildungsentscheidungen und Bildungsrenditen</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, die in BA-Verfahren erfasst werden</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>Maßnahmeteilnahmehistoriken (MTH):</b> Informationen zu Teilnahmen an Maßnahmen der aktiven Arbeitsmarktförderung (keine Maßnahmen der zugelassenen kommunalen Träger im SGB II)</p> <p><b>Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II (XMTH):</b> Informationen zu Teilnahmen an Maßnahmen zur Eingliederung, die über den Übermittlungsstandard XSozial-BA-SGB II von zugelassenen kommunalen Trägern an die BA gemeldet werden.</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>Erhebungsdaten</u>          Studienanfänger zum Studienjahr 2010/2011 an deutschen Hochschulen, unabhängig von Nationalität und Deutschkenntnissen</p> <p><u>Administrative Personendaten</u>          Sozialversicherungspflichtig Beschäftigte (ab 1999 auch geringfügig Beschäftigte), Leistungsempfänger, Arbeitssuchende, Maßnahmeteilnehmerinnen und -teilnehmer, Beschäftigungsbetriebe</p>
Fallzahlen	<p><u>Erhebungsdaten</u></p> <p>Welle 1: 17.910 Teilnehmerinnen und Teilnehmer          Welle 2: 12.273 Teilnehmerinnen und Teilnehmer          Welle 3: 13.113 Teilnehmerinnen und Teilnehmer          Welle 4: 11.202 Teilnehmerinnen und Teilnehmer          Welle 5: 13.866 Teilnehmerinnen und Teilnehmer (davon verknüpft: 9.619 Personen<sup>5</sup>)          Welle 6: 10.183 Teilnehmerinnen und Teilnehmer          Welle 7: 9.611 Teilnehmerinnen und Teilnehmer (davon zusätzlich verknüpft: 651)          Welle 8: 8.629 Teilnehmerinnen und Teilnehmer          Welle 9: 10.096 Teilnehmerinnen und Teilnehmer (davon zusätzlich verknüpft: 198)          Welle 10: 9.090 Teilnehmerinnen und Teilnehmer (davon zusätzlich verknüpft: 0)          Welle 11: 7.020 Teilnehmerinnen und Teilnehmer          Welle 12: 8.918 Teilnehmerinnen und Teilnehmer (davon zusätzlich verknüpft: 0)          Welle 13: 7.293 Teilnehmerinnen und Teilnehmer          Welle 14: 5.161 Teilnehmerinnen und Teilnehmer</p> <p><u>Administrative Daten</u>          10.468 Personen          154.589 Originalsätze          202.571 überschneidungsfreie Sätze (nach Episodensplitting)</p>
Zeitraum	<p><u>Erhebungsdaten</u></p> <p>Welle 1: 2010/2011          Welle 2: Herbst 2011</p>

<sup>5</sup> Die Frage nach der Zustimmung zur Verknüpfung der BA-Daten mit den Befragungsdaten ist nur in den CATI-Befragungen ab Welle 5 enthalten.

	<p>Welle 3: Sommer 2012  Welle 4: Herbst 2012  Welle 5: Sommer 2013  Welle 6: Herbst 2013  Welle 7: Sommer 2014  Welle 8: Herbst 2014  Welle 9: 2015  Welle 10: Sommer 2016  Welle 11: Herbst 2016  Welle 12: 2017  Welle 13: Sommer 2018  Welle 14: Herbst 2018</p> <p><u>Administrative Daten</u>  Der Zeitraum hängt von der Datenquelle ab.  BeH: 01.01.1975 - 31.12.2017 (2014: 36-Monats-Datei  2015 und 2016: 18-Monats-Datei, 2017: 6-Monats-Datei)  LeH: 01.01.1975 - 31.12.2017  ASU: 01.01.1997 - 31.12.2017  LHG: 01.01.2005 - 31.12.2017  XASU: 01.01.2005 - 31.12.2017  MTH: 01.01.2000 - 31.12.2017  XMTM: 01.01.2005 - 31.12.2017</p>
Zeitbezug	<p><u>Erhebungsdaten</u>  Befragungszeitpunkt; Spell- bzw. Ereignisdaten retrospektiv bei Erstbefragung, anschließend jeweils seit letzter Befragung</p> <p><u>Administrative Personendaten</u>  Tagesgenaue Erwerbsbiografien</p> <p><u>Administrative Betriebsdaten</u>  Stichtag 30. Juni</p>
Regionale Gliederung	<p><u>Administrative Daten</u>  Bundesländer, Kreise</p>
Gebietsstand	<p><u>Administrative Daten</u>  Gebietsstand bereinigt zum 31.12.2017</p>
Erhebungsdesign	<p><u>Erhebungsdaten</u></p> <p>Welle 1: PAPI Rekrutierung und Kompetenztestung (Lesen/Lesegeschwindigkeit/Mathematik/Metakognition) und Telefoninterview (CATI)  Welle 2: Befragung CAWI  Welle 3: Wiederholungsbefragung, CATI biografisches Interview  Welle 4: Wiederholungsbefragung, CAWI  Welle 5: Wiederholungsbefragung, CATI biografisches Interview und PAPI/PAPI(E-PEN)/CBA/CBWA Kompetenztestung (Kognitive Grundfähigkeiten/Naturwissenschaften/IKT-Kenntnisse/Metakognition )  Welle 6: Wiederholungsbefragung, CAWI</p>

	<p>Welle 7: Wiederholungsbefragung, CATI biografisches Interview mit allen Studierenden außer dem Lehramts-Oversample (Substichprobe A), PAPI Kompetenztestung (Business Administration und Economics /Metakognition) innerhalb eines CAPI mit Teilnehmern, die ein wirtschaftswissenschaftliches Fach studieren oder ein solches Studium erfolgreich abgeschlossen haben (Substichprobe B)</p> <p>Welle 8: Wiederholungsbefragung, CAWI</p> <p>Welle 9: Wiederholungsbefragung, CATI biografisches Interview</p> <p>Welle 10: Wiederholungsbefragung, CATI biografisches Interview</p> <p>Welle 11: Wiederholungsbefragung, CAWI</p> <p>Welle 12: Wiederholungsbefragung, CATI biografisches Interview und CBA/CBWA Kompetenztestung (Lesen/Mathematik/Metakognition oder Lesen/Englisch-Lesekompetenz/Metakognition oder Mathematik/Englisch-Lesekompetenz/Metakognition)</p> <p>Welle 13: Wiederholungsbefragung, CATI biografisches Interview</p> <p>Welle 14: Wiederholungsbefragung, CAWI</p> <p><u>Administrative Daten</u></p> <p>Identifikation der Befragten in den Integrierten Erwerbsbiografien des IAB</p>
Beteiligte Institutionen	<p><u>Erhebungsdaten</u></p> <p>Auftraggeber: Leibniz-Institut für Bildungsverläufe e.V. (LifBi) an der Otto-Friedrich-Universität Bamberg</p> <p>Durchführung der Befragung: infas Institut für angewandte Sozialwissenschaft GmbH und DZHW Deutsches Zentrum für Hochschul- und Wissenschaftsforschung</p> <p><u>Administrative Daten</u></p> <p>Sozialversicherungsträger, Bundesagentur für Arbeit, kommunale Träger</p>
Frequenz der Datensammlung	<p><u>Erhebungsdaten</u></p> <p>ca. halbjährlich bis jährlich</p> <p><u>Administrative Daten</u></p> <p>Laufend</p>
Dateiformat/-größe	<p><u>Erhebungsdaten</u></p> <p>Stata; 32 Einzeldatensätze zw. 198 KB und 147,23 MB</p> <p><u>Administrative Daten</u></p> <p>Stata; Personendatensatz: 15,11 MB; Betriebsdatensatz: 5,4 MB</p>
Dateiorganisation	<p><u>Erhebungsdaten</u></p> <p>Die Daten sind in mehreren Dateien abgelegt, die im Quer- oder Längsschnitt (wellen- bzw. episodenspezifisch) angelegte Personendaten enthalten.</p> <p><u>Administrative Daten</u></p> <p>Die Daten werden in zwei Dateien abgelegt. Die eine Datei enthält personenbezogene und die andere Datei betriebsbezogene Informationen. Auf begründeten Antrag werden weitere Dateien mit zusätzlichen Betriebsinformationen bereitgestellt.</p> <p><u>Verknüpfungsdaten</u></p> <p>Daneben gibt es eine Datei, die die Identifikationsvariable sowie Informationen zum Matching-Erfolg beinhaltet.</p>
Datenzugang	<p>Gastaufenthalt am IAB-FDZ, Datenfernverarbeitung beim IAB-FDZ nach vorangegangenem Gastaufenthalt</p>

Anonymisierungsgrad	<u>Erhebungsdaten</u> Faktisch anonym <u>Administrative Daten</u> Schwach anonymisiert
Sensible Merkmale	<u>Administrative Daten</u> Geburtsmonat (gebmon), 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	<u>Daten</u> „Die Datengrundlage dieses Beitrags bilden die NEPS-SC5-Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC5-ADIAB). Der Datenzugang erfolgt über einen Gastaufenthalt am Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ) und/oder mittels kontrollierter Datenfernverarbeitung beim FDZ.“ DOI: 10.5164/IAB.NEPS-SC5-ADIAB7518.de.en.v1 <u>Datendokumentation</u> Bachbauer, Nadine; Wolf, Clara (2020): NEPS-SC5-Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC5-ADIAB 7518). FDZ-Datenreport, 10/2020 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.2010.de.v1
Datensatzversion	NEPS-SC5 Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC5-ADIAB 7518); DOI: 10.5164/IAB.NEPS-SC5-ADIAB7518.de.en.v1

## 1.4 Merkmalsliste der administrativen Daten

Die Merkmalsübersicht in Tabelle 4 listet die Variablennamen und die längeren Merkmalsbezeichnungen der administrativen Daten auf. Seitens der Betriebsdaten sind hier wie in Kapitel 6 nur die Merkmale der Basisbetriebsdatei aufgeführt. Die weiteren Betriebsmerkmale sowie die Erweiterungsmodule werden ausführlich in Schmucker et al. (2018) beschrieben. Weiterhin bietet Tabelle 4 eine Übersicht darüber, welche Merkmale für welche Datenquellen gefüllt sind, wobei Tabelle 3 aufzeigt, wie die Füllgrade zu interpretieren sind.

Tabelle 3 Füllgrad der Merkmale

<b>h</b>	Merkmal ist für die Quelle gefüllt. Füllgrad immer > 0,85
<b>m</b>	Merkmal ist für die Quelle gefüllt. Niedriger bzw. wechselnder Füllgrad, siehe Merkmalsbeschreibung und Häufigkeitsauszählung
<b>l</b>	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 Variable 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 Personen-gruppe 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 bzw. XMTH die Maßnahmenkategorie. Nicht bei jedem Merkmal sind diese Unterschiede direkt aus dem Variablennamen ersichtlich.

Tabelle 4 Merkmalsliste der administrativen Daten mit Füllgraden

Merkmalsliste	Modul	Seite	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
<b>Identifikatoren</b>		<b>45</b>							
Personen-ID (ID_t)		45	h	h	h	h	h	h	h
Betriebs-ID (betnr)	*	45	h	l	l	l	l	l	l
<b>Generierte technische Merkmale</b>		<b>46</b>							
Satzzähler pro Konto (spell)		46	h	h	h	h	h	h	h
Quelle des Satzes (quelle)		46	h	h	h	h	h	h	h
Jahr (jahr)	*	47	l	l	l	l	l	l	l
<b>Gültigkeitszeitraum</b>		<b>47</b>							
Beginndatum Originalbeobachtung (begorig)		47	h	h	h	h	h	h	h
Enddatum Originalbeobachtung (endorig)		47	h	h	h	h	h	h	h
Beginndatum der Episode (begepi)		48	h	h	h	h	h	h	h
Enddatum der Episode (endepi)		48	h	h	h	h	h	h	h
<b>Informationen zur Person</b>		<b>48</b>							
Geschlecht (frau)		48	h	h	h	h	h	h	h
Geburtsjahr (gebjahr)		49	h	h	h	h	h	h	h
Geburtsmonat (gebmon)		49	h	h	h	h	h	h	h
Staatsangehörigkeit (nation)		49	h	h	h	h	h	h	h
Staatsangehörigkeit vergrößert (nation_gr)		50	h	h	h	h	h	h	h
Familienstand (famst)		50	l	h	h	h	h	m	h
Kinderzahl (kind)		50	l	l	l	m	m	m	m
Ausbildung (ausbildung)		51	m	l	l	h	m	h	m

<b>Merkmalsliste</b>	Modul	Seite	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Ausbildung (imputiert) (ausbildung_imp)		53	h	l	l	l	l	l	l
Schulausbildung (schule)		54	m	l	l	h	m	h	m
<b>Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche</b>		<b>55</b>							
Tagesentgelt / täglicher Leistungssatz (tentgelt)		55	h	h	l	l	l	l	l
Beruf – ausgeübte / letzte Tätigkeit (KldB 1988) (beruf)Beruf – ausgeübte / letzte Tätigkeit (KldB 1988) (beruf)		56	h	l	l	m	l	m	l
Berufsgruppe – ausgeübte / letzte Tätigkeit (KldB 2010) (beruf2010_3)		56	h	m	l	m	m	m	m
Berufsuntergruppe – ausgeübte / letzte Tätigkeit (KldB 2010) (beruf2010_4)		57	h	m	l	m	m	m	m
Anforderungsniveau – ausgeübte / letzte Tätigkeit (KldB 2010) (niveau)		58	h	m	l	m	m	m	m
Teilzeit (teilzeit)		59	h	l	l	l	l	l	l
Stellung im Beruf und Arbeitszeit (stib)		59	h	l	l	l	l	l	l
Erwerbsstatus (erwstat)		60	h	h	h	h	h	h	h
Gleitzone (gleitz)		61	h	l	l	l	l	l	l
Leiharbeit (leih)		62	h	l	l	l	l	l	l
Befristung (befrist)		62	h	l	l	l	l	l	l
Grund Abgabe Beschäftigungsmeldung / Grund Ende Leistungsbezug (grund)		62	h	h	m	h	m	l	l
Erwerbsstatus vor Arbeitsuche (estatvor)		64	l	l	m	h	m	l	l
Status nach Arbeitsuche (estatnach)		64	l	l	m	l	l	l	l
Profillage (profil)		64	l	l	m	l	m	l	l
Art der Kündigung der letzten Tätigkeit (art_kuend)		65	l	l	m	l	m	l	l
Arbeitszeit des Stellengesuchs (arbeit)		65	l	l	m	l	m	l	l
Restanspruch / geplante Dauer (restanspruch)		65	l	h	l	l	h	l	l
Trägerart (traeger)		66	l	l	h	h	h	h	h

<b>Merkmalsliste</b>	Modul	Seite	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Beginndatum der Arbeitslosigkeit (alo_beg)		66	<b>m</b>						
Dauer der Arbeitslosigkeit (alo_dau)		67	<b>h</b>						
<b>Ortsangaben</b>		<b>67</b>							
Wohnort – Kreis (wo_kreis)		67	<b>h</b>						
Wohnort – Bundesland (wo_bula)		68	<b>h</b>						
Wohnort – Arbeitsagentur (wo_aa)		68	<b>h</b>						
Wohnort – Regionaldirektion (wo_rd)		69	<b>h</b>						
<b>Betriebsmerkmale</b>		<b>69</b>							
WZ73 3-Steller (w73_3)	*	69	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
WZ93 5-Steller (w93_5)	*	70	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
WZ93 3-Steller (w93_3)	*	70	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
WZ03 5-Steller (w03_5)	*	71	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
WZ03 3-Steller (w03_3)	*	71	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
WZ08 5-Steller (w08_5)	*	72	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
WZ08 3-Steller (w08_3)	*	72	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
w73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen)	*	72	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
Art der Vervollständigung w73_3 (group_w73_3)	*	73	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
w93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen)	*	73	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
Art der Vervollständigung w93_3 (group_w93_3)	*	73	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
w08_3 vervollständigt durch Extrapolation/Imputation (w08_3_gen)		74	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
Art der Vervollständigung w08_3 (group_w08_3)		74	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
Jahr des 1. Auftretens der Betriebsnummer (grd_jahr)	*	74	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
Erstes Auftreten Betriebsnummer (grd_dat)	*	75	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>
Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr)	*	75	<b>h</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>	<b>l</b>

<b>Merkmalsliste</b>	Modul	Seite	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Letztes Auftreten Betriebsnummer (lzt_dat)	*	76	h	l	l	l	l	l	l
Anzahl Beschäftigte gesamt (az_ges)	*	76	h	l	l	l	l	l	l
Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz)	*	76	h	l	l	l	l	l	l
Anzahl geringfügig Beschäftigte (az_gf)	*	77	h	l	l	l	l	l	l
Mittelwert imputiertes Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)	*	77	h	l	l	l	l	l	l
Arbeitsort Kreis (ao_kreis)	*	77	h	l	l	l	l	l	l
Arbeitsort Bundesland (ao_bula)	*	78	h	l	l	l	l	l	l

## 1.5 Mengengerüst

In Tabelle 5 ist die Anzahl der Episoden nach Quelle sowie vor und nach dem Episodensplitting, welches in Kapitel 3.2 näher erläutert wird, abgebildet.

Tabelle 5 Mengengerüst

<b>Fallzahlen</b>	<b>vor Episodensplitting</b>	<b>nach Episodensplitting</b>
BeH	126.739	156.238
LeH	3.828	5.955
LHG	4.844	10.325
ASU	16.566	25.757
XASU	944	1.679
MTH	1.559	2.379
XMTH	109	238
Sätze insgesamt	154.589	202.571
Personen		10.468

## 2 Datenquellen

### 2.1 NEPS-SC5-Erhebungsdaten (NEPS-SC5)

Die NEPS-SC5-Teilstudie „Hochschulstudium und Übergang in den Beruf“ stellt eine Datengrundlage bereit, die Forschung zu Studienabbruch, sozialer Selektivität beim Hochschulzugang sowie zum Verhältnis von Hochschule und Beruf ermöglicht. Darüber hinaus können neue Herausforde-

rungen, die beispielsweise durch die Einführung der zweistufigen Studienstruktur entstehen, untersucht werden. In diesem Zusammenhang sind auch Themen wie die wachsende Forderung nach Outcome-Orientierung und der Fachkräftemangel im Bereich der Hochqualifizierten zu nennen, die mit den Daten analysiert werden können. Dazu wurden insbesondere Bildungsentscheidungen, die Entwicklung von Kompetenzen sowie die Erträge eines Studiums und der Übergang in den Beruf erfasst.

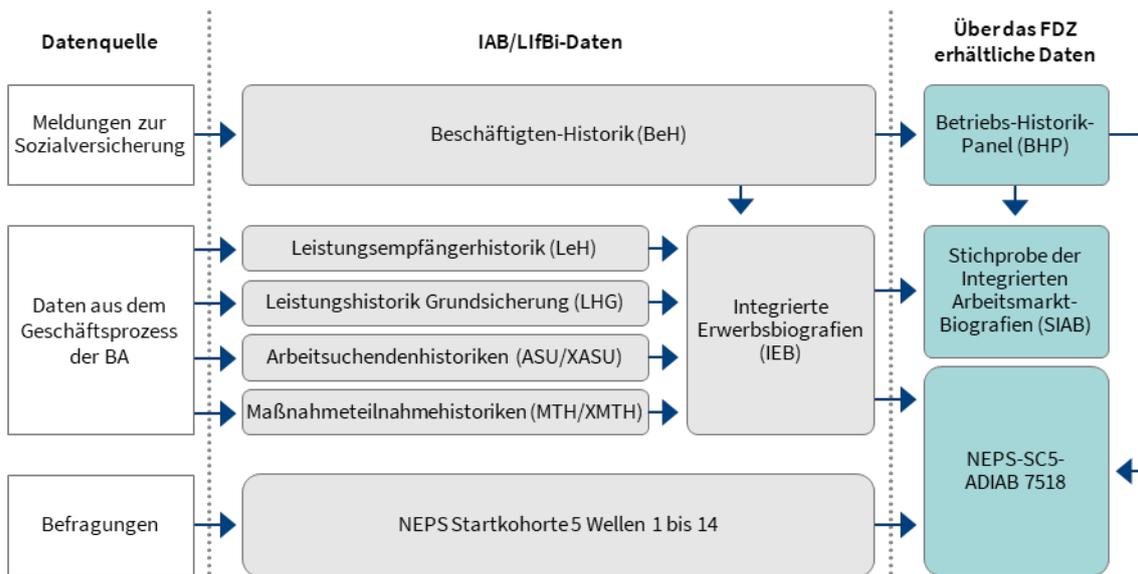
Inhaltlich sollen die erhobenen Daten es unter anderem ermöglichen,

- die Kompetenzentwicklung von Studierenden im Studienverlauf zu beobachten und darüber hinaus zu identifizieren, welche fachlichen und außerfachlichen Kompetenzen durch das Studium erlangt werden und wie der Erwerb dieser durch Lernumwelten und individuelle Lernstrategien determiniert wird;
- den Einfluss der Fachrichtung und des Hochschultyps auf den Kompetenzerwerb zu analysieren;
- die Übereinstimmung von angeeigneten Kompetenzen mit den Anforderungen des Beschäftigungssystems zu untersuchen;
- die Einflussfaktoren von Studienentscheidungen z.B. hinsichtlich Studienabbruch oder Studienfachwechsel sowie die Konsequenzen solcher Entscheidungen für den weiteren Bildungs- und Berufsverlauf zu bestimmen;
- die Determinanten von Studienerfolg sowie die Rolle von Kompetenzen und sozialen Faktoren in Zusammenhang mit Studienentscheidungen und -erfolg zu ermitteln;
- die Relevanz von erworbenen Kompetenzen für den Übergang in den Beruf und den Berufserfolg der Bedeutung von formalen Qualifikationen, sozialen Faktoren sowie sozialem und kulturellem Kapital gegenüberzustellen;
- die Unterschiede zwischen fachspezifischen und fachübergreifenden Kompetenzen im Zusammenhang mit Berufseinstieg und -erfolg herauszuarbeiten.

## 2.2 Administrative Daten aus den Integrierten Erwerbsbiografien (IEB)

Die administrativen Personendaten werden aus den Integrierten Erwerbsbiografien (IEB) des IAB gezogen und hinsichtlich Variablenauswahl und -ausprägungen entsprechend der vom FDZ erstellten Stichprobe der integrierten Arbeitsmarktbiografien (SIAB, siehe Antoni et al., 2019) aufbereitet. Die IEB 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 zu NEPS-SC5-ADIAB führen und stellt deren Beziehung zu anderen Datenprodukten des FDZ dar.

Abbildung 2 Quellen von NEPS-SC5-ADIAB 7518



### 2.3 Beschäftigtenhistorik (BeH)

Die Grundlage der Daten ist das mit Wirkung vom 01.01.1973 eingeführte (und seit dem 01.01.1991 auf Ostdeutschland ausgedehnte) integrierte Meldeverfahren zur Kranken-, Renten- und Arbeitslosenversicherung (DEÜV; früher DEVO/DÜVO; 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 (vgl. Cramer, 1985) grundsätzlich nicht in der BeH erfasst werden. Seit der Änderung des Meldeverfahrens zum 01.01.1999 werden auch geringfügig Beschäftigte und mithelfende Familienangehörige erfasst (enthalten erst ab 01.04.1999). 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 30.06. eines Jahres aggregiert werden. Die Betriebsmerkmale sind deshalb, im Unterschied zu den Personenangaben, keine zeitraumbezogenen Daten, sondern nur am 30.06. jedes Jahres exakt gültig.<sup>6</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

<sup>6</sup> Extrembeispiel: Es liegen Beschäftigungsmeldungen vom 01.01 bis 30.05.2006 vor, der Betrieb meldet jedoch Konkurs an, weshalb alle Beschäftigten entlassen werden. Dann gibt es im BHP zu diesem Betrieb für 2006 keine Informationen.

1). Nicht alle Merkmale des BHP sind standardmäßig in NEPS-SC5-ADIAB enthalten. Auf begründeten Antrag hin können einzelne weitere Merkmale des BHP<sup>7</sup> vom FDZ zugespielt werden.

---

#### Kasten 1 Beispielcode zur Verbindung von Personen- und Betriebsdaten in Stata 14

```
use NEPS-SC5-ADIAB_7518_v1.dta, clear
gen int jahr = year(begepi)
sort betnr jahr
merge m:1 betnr jahr using NEPS-SC5-ADIAB_7518_v1_bhp_basis_v1.dta
```

## 2.4 Leistungsempfängerhistorik (LeH)

Die Leistungsempfängerhistorik (LeH) erfasst Zeiträume, in denen Personen Lohnersatzleistungen der Bundesagentur für Arbeit beziehen (Rechtskreis SGB III). Die Leistungen beinhalten Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld, aber 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 01.01.1975.

## 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 (ARGEn, 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>8</sup>,

---

<sup>7</sup> Nähere Informationen zum BHP finden sich bei Schmucker et al. (2018) und unter <http://fdz.iab.de/>.

<sup>8</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.

- Zugelassene kommunale Träger, die nach der anfänglichen Experimentierklausel des § 6a SGB II auch als optierende Kommunen oder Optionskommunen bezeichnet werden – hier übernimmt die Kommune alle Aufgaben aus dem Bereich des SGB II.

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>9</sup>. Zugelassene kommunale Träger verwenden verschiedene eigene IT-Verfahren und übermitteln ihre Daten mittels des Standards XSozial-BA-SGB II an die BA. 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 01.01.2005. Die Datenquelle ist jedoch bis Anfang 2007 lückenhaft (siehe Kapitel 5.5).

## 2.6 Arbeitssuchendenhistorik (ASU und XASU)

Daten über Arbeitssuchende werden in der Arbeitssuchendenhistorik (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. Der linke Datenrand der ASU ist der 01.01.1997, der der XASU ist der 01.01.2005.

## 2.7 Maßnahmeteilnahmehistoriken (MTH und XMTH)

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. In der Quelle XMTH sind ab 2005 Maßnahmen der zugelassenen kommunalen Träger bzw. optierenden Kommunen geführt. Der linke Datenrand der MTH ist der 01.01.2000, der der XMTH ist der 01.01.2005.

## 2.8 Veränderungen zur SIAB 7514

### 2.8.1 Korrektur der Fehlbuchungen in der Gemeinde Lahn

Für die Jahre 1975-1977 gab es in der Quelle BeH eine historische Fehlzuzuordnung der Betriebe und Beschäftigten der Gemeinde Lahn (Niedersachsen). Diese Fehlzuzuordnungen wurden soweit wie möglich korrigiert.

---

<sup>9</sup> 2014 wurde A2LL sukzessive durch ALLEGRO als neues IT-Verfahren für die Grundsicherung im SGB II in gemeinsamen Einrichtungen ersetzt.

## 2.8.2 Aufnahme der Maßnahmeteilnahmehistorik aus XSozial\_BA-SGB II (XMTH)

Die SIAB 7517 enthält anders als die Vorgängerversion auch Informationen aus den ebenfalls in den IEB enthaltenen Maßnahmeteilnahmehistoriken, die über den Übermittlungsstandard XSozial-BA-SGB II von zugelassenen kommunalen Trägern an die BA gemeldet werden (XMTH). Die darin enthaltenen Maßnahmekategorien werden jedoch nur in gruppierter Form angeboten.

## 2.8.3 Merkmalskanon

Die SIAB 7517 bietet gegenüber der SIAB 7514 eine Reihe von zusätzlichen Merkmalen. Das in der vorherigen Version entfernte Merkmal „Stellung im Beruf und Arbeitszeit“ (stib) ist in der SIAB 7517 wieder enthalten. Es gilt jedoch die Einschränkung, dass das Merkmal nur für Meldungen gefüllt ist, die vor der Einführung des neuen Tätigkeitsschlüssels abgegeben wurden (siehe Abschnitt 3.1.3). Neu aufgenommen wurde zudem das Merkmal „Ausbildung (imputiert)“ (ausbildung\_imp, siehe Abschnitt 6.4.9). Diese Variable enthält für Beobachtungen mit ursprünglich fehlender Angabe zum beruflichen Ausbildungsniveau imputierte Angaben. Aufgrund zunehmender Anfragen enthält die SIAB 7517 nun auch das sensible Merkmal „Geburtsmonat“ (gebmon, siehe Abschnitt 6.4.3). Zudem wurden zeitkonsistente Wirtschaftszweige der Klassifikation WZ08 durch Extrapolation und Imputation generiert. Weitere Informationen dazu finden sich in den Abschnitten 6.7.12 und 6.7.13. Tabelle 6 gibt einen Überblick über die Änderungen im Merkmalskanon über alle Quellen hinweg.

Tabelle 6: Neue Merkmale in der SIAB 7517

<b>Merkmal</b>	<b>Erläuterung</b>
Ausbildung imputiert (ausbildung_imp)	Neu aufgenommen, jedoch mit weniger Ausprägungen als die Variable „Ausbildung“ (ausbildung)
Geburtsmonat (gebmon)	Neu als sensibles Merkmal aufgenommen
Stellung im Beruf und Arbeitszeit (stib)	Wieder aufgenommen, jedoch nur für Meldungen gefüllt, die vor der Einführung des neuen Tätigkeitsschlüssels abgegeben wurden
w08_3 vervollständigt durch Extrapolation/Imputation (w08_3_gen)	Neu aufgenommen in Basisbetriebsdatei
Art der Vervollständigung w08_3 (group_w08_3)	Neu aufgenommen in Basisbetriebsdatei

# 3 Datenaufbereitung und Stichprobenziehung

## 3.1 Bereinigungen und Aufbereitungen

### 3.1.1 NEPS-SC5-Erhebungsdaten

Detaillierte Informationen zur Stichprobenziehung, zur Datenerhebung, zu den eingesetzten Instrumenten sowie zur Struktur und Aufbereitung der NEPS-SC5-Erhebungsdaten finden sich auf der NEPS-Website. Neben einer Reihe von Reports und Tools für den Umgang mit den Daten steht dort auch das sogenannte „Semantische Datenstruktur-File“ zur Verfügung, welches einen ersten Einblick in die Datensätze der SC5 inklusive Variablennamen, Variablen- und Wertelabels (jedoch ohne Beobachtungen) ermöglicht.<sup>10</sup>

### 3.1.2 Administrative Daten aus den Integrierten Erwerbsbiographien (IEB)

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:

- Sätze, bei denen das Alter an ihrem Ende kleiner 13 oder zu Beginn größer 75 ist, werden gelöscht.
- Sätze, bei denen das Enddatum vor dem Beginndatum liegt, werden gelöscht.
- Widersprüchliche Angaben zu Geschlecht oder Geburtsdatum innerhalb eines Kontos werden bereinigt.
- Sätze 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.

### 3.1.3 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 in 2011 neu eingeführt.

---

<sup>10</sup> Siehe <https://www.neps-data.de>.

- 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 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 31.12.2017.

#### 3.1.4 Leistungsempfängerhistorik (LeH)

- Sätze ohne gültiges Beginndatum wurden ausgeschlossen.
- Liegt das Enddatum bei Bezug von Arbeitslosenhilfe einen Tag vor dem Beginndatum, so wurde das Enddatum um ein Jahr hochgesetzt und der Spell nicht gelöscht.
- Zwischen 2004 und 2006 wurde das zugrundeliegende 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.
- Es wird eine Gebietsstandbereinigung analog zur BeH durchgeführt.

#### 3.1.5 Leistungshistorik Grundsicherung (LHG)

- Sätze ohne BA-Kundennummer werden gelöscht.
- Stornierte Sätze 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 den 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,
  - bei Veränderungen in der Zusammensetzung einer Bedarfsgemeinschaft (z.B. durch Zu- / Abgänge),
  - bei Merkmalsänderungen des Bedarfsgemeinschafts-Kundinnen und -Kunden sowie
  - bei Beginn und Ende eines Sanktionssachverhalts für Sätze ab dem 01.04.2006. 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 der Trägerneuorganisation 2012 wurden soweit möglich bereinigt (siehe Kapitel 5.5).
- Es wird eine Gebietsstandbereinigung analog zur BeH durchgeführt.

### 3.1.6 Arbeitssuchendenhistorik (ASU und XASU)

- Sätze mit Enddatum vor dem 01.01.1995 wurden ausgeschlossen.
- 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 ASU und XASU gefüllt sind, beziehen sich immer auf den Beginn des Satzes.
- Ein neuer ASU-Satz wird angelegt, sobald ein Statuswechsel (z.B. von arbeitssuchend zu arbeitslos) erfolgt. Dies gilt auch, wenn sich die Trägerart (Agentur für Arbeit, ARGE, gemeinsame Einrichtung, zugelassener kommunaler Träger, getrennte Aufgabenwahrnehmung) ändert.
- In der XASU werden überschneidungsfreie Zeiträume bezogen auf eine einzelne Person abgebildet. Bei Änderungen folgender Merkmale wird für die XASU jeweils ein neuer Satz gebildet:
  - Wechsel des Arbeitsuche-Status
  - Wechsel der Verfügbarkeit
  - Wechsel des SGB-II-Trägers
  - Wechsel des Wohnortes
- Eine Gebietsstandbereinigung wird analog zur BeH durchgeführt.

### 3.1.7 Maßnahmeteilnahmehistoriken (MTH)

- Sätze, die mehr als ein Jahr nach Maßnahmeende erstellt wurden, werden gelöscht, falls ein anderer Satz existiert, der innerhalb eines Jahres nach Abschluss der Maßnahme erstellt wurde.
- Es wird lediglich der aktuellste Satz 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 Maßnahmenteilnahmen werden nicht aufgenommen. Hierzu zählen Unterstützung der Beratung und Vermittlung, Mobilitätshilfen und reine Reha-Maßnahmen.
- Eine Gebietsstandbereinigung wird analog zur BeH durchgeführt.

### 3.1.8 Maßnahmeteilnahmehistoriken aus XSozial (XMTH)

- Für die XMTH besteht die besondere Herausforderung darin, technisch oder organisatorisch verursachte Mehrfachmeldungen für eine Maßnahmeteilnahme zu identifizieren und zu behandeln. Die Ursachen von diesen Doppelmeldungen sind zum einen, eine bis April 2009 fehlende Identifikationsnummer für Maßnahmeteilnahmen bzw. unsachgemäße Handhabung

der eingeführten Förder-ID, zum anderen Trägernummernwechsel im Zusammenhang mit Fusionen oder anderen Umorganisationen von Grundsicherungsträgern mit der Neuorganisation der Arbeitsagenturen 2012/2013. Zudem kam es zu trägerspezifischen Problemen bei Versions- oder Produktwechseln der kommunalen Software oder bei Versionsänderungen des Standards XSozial. Daraus resultieren zeitweise in beträchtlichem Umfang Meldungen für Maßnahmeteilnahmen der gleichen Förderungsart, die sich pro Teilnehmendem teilweise oder vollständig zeitlich überlappen. Echte Dubletten werden identifiziert und aussortiert, da sie mit hoher Wahrscheinlichkeit technische Dubletten sind. Versteckte Dubletten sind ebenfalls identifizierbar und werden aussortiert. Dabei gilt im Prinzip immer die Regel, dass der aktuellste Informationsstand behalten wird.

- Daneben werden Meldungen mit einem Beginndatum vor 2005 ausgeschlossen.
- Überlappende und unmittelbar aneinander angrenzende Meldungen der gleichen Maßnahmengattung (aus XSozial) werden zu einer Beobachtung zusammengefasst. Maßnahmenspezifische Merkmale der zusammengefassten Zeiträume werden auf Systemmissings gesetzt.
- Anschließend werden die weiteren Merkmale zusammengestellt und berechnet. Dabei wird auf konsolidierte Personendaten aus XSozial-Historiken und LHG zurückgegriffen.
- Abschließend werden folgende Maßnahmen ausgeschlossen:
  - Einmalleistungen (wie UBV / Mobi- / Vermittlungsbudget / LES) und
  - spezifische Reha-Maßnahmen
- Eine Gebietsstandsberichtigung wird analog zur BeH durchgeführt.

### 3.1.9 SGB-II-Anonymisierung

Um das Deanonymisierungsrisiko zu verringern, wird in den administrativen Daten 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.

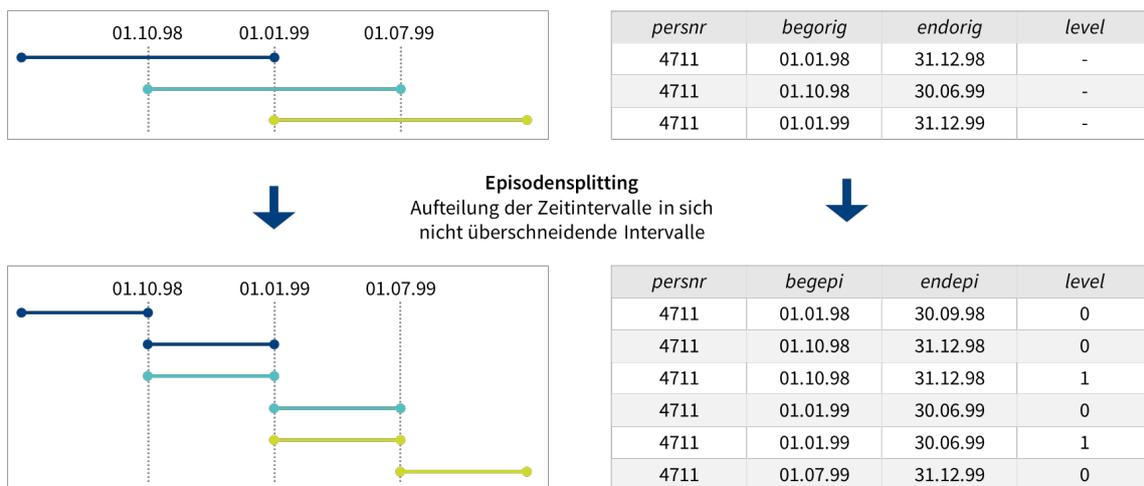
Um eine indirekte Bestimmung des exakten Geburtstags zu verhindern, wird daher das folgende Verfahren angewendet. Sätze, die am 18., 25. oder 65. Geburtstag geteilt wurden, werden zu einem einzelnen Satz zusammengefügt. Hierdurch geht außer dem exakten Geburtstag keine weitere Information verloren. Bei 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. 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 zum Beginn und Ende des Originalsatzes (*begorig* und *endorig*) bleiben erhalten, die Variablen *Beginn* und *Ende* der Episode (*begepi* und *endept*) 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 *endept*) festgestellt werden. Um die Originaldaten ohne Episodensplitting wiederherzustellen 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:

**Kasten 2 Beispielcode zum Erstellen weiterer Satzzähler in Stata 14**

```
bysort ID_t begepi quelle (spell): gen byte level1 = _n-1
bysort ID_t begepi (spell): gen byte level2 = _n-1
```

### 3.3 Fehlende Werte

In den Daten gilt die in Tabelle 7 beschriebene Codierung von fehlenden Werten.

Tabelle 7 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 in Kapitel 1.4) oder für einen gewissen Zeitraum nicht gefüllt.

## 4 Datenverknüpfung

### 4.1 Informierte Zustimmung der Befragten zur Datenverknüpfung

Ausgangspunkt der Verknüpfung der Erhebungs- und Prozessdaten sind die Befragten der NEPS-SC5. Diese wurden im Rahmen des Interviews um Erlaubnis gebeten, ihre Erhebungsdaten für Forschungszwecke mit den über sie am IAB vorhandenen administrativen Daten verknüpfen zu dürfen. Der Fragebogenausschnitt mit der Zustimmungserklärung lautet:

*„Um das Interview im Folgenden möglichst kurz zu halten, würden wir gerne bei der Auswertung der Befragung Auszüge aus Daten einbeziehen, die beim Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit in Nürnberg vorliegen. Dabei handelt es sich zum Beispiel um Informationen zu Beschäftigungsverhältnissen, Phasen der Arbeitslosigkeit und Teilnahmen an Maßnahmen während einer Arbeitslosigkeit sowie um betriebliche Merkmale. Zum Zweck der Zuspiegelung dieser Daten an die Interviewdaten möchte ich Sie herzlich um Ihr Einverständnis bitten. Werden diese Informationen ausgewertet, so ist absolut sichergestellt, dass alle datenschutzrechtlichen Bestimmungen strengstens eingehalten werden und keine Weitergabe der Daten an Dritte erfolgen wird.*

*Ihr Einverständnis ist selbstverständlich freiwillig. Sie können es auch jederzeit bei der Ansprechpartnerin bei infas widerrufen. Sind Sie mit der Zuspiegelung von Daten des Instituts für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit einverstanden?“*

Die Befragten mussten der obenstehenden Frage zustimmen, um Teil der Stichprobe für die Verknüpfung zu werden. Die Frage zum Einverständnis zur Verknüpfung wurde innerhalb der CATI-Befragungen der Wellen 5 bis einschließlich 13 gestellt, wobei Personen diese Frage nur ein weiteres Mal erhielten, wenn sie der Verknüpfung zuvor nicht zugestimmt hatten.

### 4.2 Herkunft der Verknüpfungsmerkmale

Für die Verknüpfung der Erhebungsdaten der in NEPS-SC5 befragten und zuspielbereiten Personen mit deren administrativen IAB-Daten wurden folgende personenbezogenen Merkmale aus beiden Datenquellen herangezogen:

- Vor- und Nachname
- Geburtsdatum
- Geschlecht
- Wohnadresse (Postleitzahl, Ort, Straße, Hausnummer)

Seitens der NEPS-Erhebung wurden diese Merkmale vom Institut für angewandte Sozialwissenschaft (infas) aus den Befragten-Kontaktdaten an das IAB übermittelt. Bei den enthaltenen Personen handelt es sich ausschließlich um Befragte, die der Verknüpfung explizit zugestimmt haben. Personen, die zwischenzeitlich ihre Panelbereitschaft zurückgezogen haben und zuvor der Verknüpfung zugestimmt hatten, wurden bei der Verknüpfung nicht berücksichtigt. Dadurch besteht eine Differenz zwischen der Anzahl der Personen, die im Rahmen der NEPS-Befragung der Verknüpfung zugestimmt haben und der Anzahl der Personen, die für das Linkage zur Verfügung standen. Zur Vorbereitung für die Datenverknüpfung wurden exakt identische Zeilen pro Person für die unterschiedlichen Erhebungswellen gelöscht. Bei Änderungen bzw. Unterschieden in der Schreibweise im Namen oder der Adresse wurden multiple Einträge pro Person behalten, da diese verknüpfungsrelevant sein konnten.

Zur Identifizierung der befragten Personen in den administrativen Daten des IAB wurde ein Auszug aus den zentralen Adressdateien des Data Warehouse der Bundesagentur für Arbeit verwendet. Dabei wurde vorab eine Beschränkung der Adressdaten auf die in NEPS-SC5 befragten Geburtsjahrgänge und von den Sampling Points abgeleiteten Postleitzahl-Bereiche vorgenommen.

In beiden Datenquellen können pro Person mehrere unterschiedliche Einträge vorhanden sein. Unterschiede können z.B. durch Umzüge, Namensänderungen oder durch uneinheitliche Erfassung eigentlich identischer Angaben entstehen. Um einen hohen Verknüpfungserfolg zu ermöglichen, werden alle unterschiedlichen Angaben zu einzelnen Personen in den Verknüpfungsprozess einbezogen.

### 4.3 Bereinigung und Standardisierung personenbezogener Merkmale

Die oben genannten personenidentifizierenden Merkmale der NEPS und administrativen Daten wurden Bereinigungsroutrinen unterzogen, um eine möglichst hohe Übereinstimmung zwischen beiden Datenquellen zu erreichen. Für dieses sog. Preprocessing wurden etablierte Verfahren eingesetzt, die am Standort des German Record Linkage Center (GRLC)<sup>11</sup> am IAB-FDZ für die Besonderheiten des deutschen Kontexts verfeinert wurden. Siehe Christen (2012) für einen umfassenden Überblick über alle in Kapitel 4.3 bis 4.5 beschriebenen Methoden und z. B. Schild/Antoni (2014) für weitere Details zu deren Anpassung an den deutschen Kontext.

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 jeweils in einem Feld ohne Leerzeichen zusammengefasst. Neben diesen grundlegenden Bereinigungen wurden variablen-spezifische Routinen zur Standardisierung der Daten durchgeführt. So wurde die Hausnummer aus einem kombinierten Straße-/Hausnummernfeld extrahiert und alle übrigen Adresszusätze (z.B. Stockwerk, Appartement-Nr.) gelöscht. Bei Straßennamen wurden häufig vorkommende

---

<sup>11</sup> Siehe Antoni/Schnell (2019) für mehr Information zum GRLC.

Muster und Abkürzungen standardisiert. Bei Ortsnamen wurden Anhänge auf eine einheitliche Form gebracht (z. B. bei Ortschaften mit zusätzlicher Information wie Fluss oder Region) bzw. entfernt (z.B. bei Berlin-Kreuzberg).

#### 4.4 Exakter Abgleich der Verknüpfungsmerkmale

Nach Bereinigung und Standardisierung der Namens- und Adresskomponenten erfolgte ein exakter Abgleich der Felder zwischen den Einträgen der beiden Datenquellen. Zunächst wurden alle Datenzeilen auf eine exakte Übereinstimmung aller zur Verfügung stehenden Merkmale (Vor- und Nachname, Geburtsdatum, Geschlecht, Straße, Hausnummer, PLZ, Ort) hin geprüft.<sup>12</sup> Anschließend wurde untersucht, wie viele weitere Personen sich verknüpfen lassen, wenn eine Abweichung auf einzelnen Adressfeldern oder dem Geburtsdatum toleriert wird (während die Vor- und Nachnamensfelder weiterhin exakt übereinstimmen müssen). Die Toleranz wurde dabei so gewählt, dass ein sehr geringes Risiko bestand, durch dieses Verfahren Einträge zu zwei unterschiedlichen Personen zu verknüpfen (sog. falsch-positive Matches). Es wurden daher zunächst Abweichungen bei einem der besonders fehleranfälligen Merkmale Hausnummer, Postleitzahl, Wohnort, Geburtstag/-monat, Straße oder Geburtsjahr zugelassen.

#### 4.5 Fehlertoleranter Abgleich der Verknüpfungsmerkmale

Für einen Teil der Personen konnte durch keinen dieser exakten Abgleiche eine Entsprechung in den administrativen Adressdaten ermittelt werden. Um diese Personen mit den administrativen Daten zu verknüpfen, wurden fehlertolerante Linkage-Verfahren verwendet, welche Ähnlichkeitsmaße der Namens- und Adresskomponenten berechnen und die Zeilenpaare mit der höchsten Qualität auswählen. Dabei wurde für jedes abzugleichende Variablenpaar ein Ähnlichkeitsmaß berechnet und diese in einem zusammenfassenden Qualitätsindex aufsummiert. Die Berechnungen wurden mit der Software Merge Tool Box (siehe Schnell et al., 2004) durchgeführt. Für Vor- und Nachname sowie Ort und Straße<sup>13</sup> wurde jeweils mittels der Jaro-Winkler-Metrik eine Stringähnlichkeit berechnet. Bezüglich der Felder Hausnummer, Postleitzahl, Geburtsdatum (zerlegt in die Komponenten Tag, Monat und Jahr) und Geschlecht wurde ein exakter Vergleich durchgeführt, da Stringähnlichkeitsmaße hier nicht sinnvoll sind.

Beim fehlertoleranten Linkage wurde ein Blocking der Beobachtungen auf Ebene der dreistelligen Postleitzahlgebiete durchgeführt. Dies hat den Vorteil, dass die Anzahl der erforderlichen Vergleiche nicht dem kartesischen Produkt der in den beiden Adressbeständen vorhandenen Zeilen entspricht, sondern auf die Beobachtungen mit identischer dreistelliger PLZ beschränkt bleibt. Zusätzlich wird hierdurch das Risiko der Verknüpfung unterschiedlicher Personen mit sehr ähnlichen Identifikationsmerkmalen verringert. Auf der anderen Seite kann es jedoch in geringem Umfang vorkommen, dass für Personen dadurch keine Entsprechung in den administrativen Daten gefunden werden kann – etwa durch fehlende bzw. fehlerhafte Angaben zur Postleitzahl oder weil ein

---

<sup>12</sup> Eine Übereinstimmung liegt bezüglich der bereinigten und standardisierten Namens- und Adressfelder vor, nicht bezüglich der unbereinigten Merkmale. Es werden im Rahmen des Preprocessings keine Bereinigungen vorgenommen, die einen Informationsverlust in der Weise nach sich ziehen würden, dass ungleiche Personen einander zugeordnet werden könnten.

<sup>13</sup> Als Robustheitstest wurde die Stringähnlichkeit des Straßennamens mit N-Grammen berechnet. Da mit Hilfe des Jaro-Winkler-Algorithmus zusätzliche Matches gefunden werden konnten und die restlichen Zuweisungen identisch sind, wird Jaro-Winkler verwendet.

Umzug in ein anderes Postleitzahlgebiet nur in einer der beiden Datenquellen registriert ist. Deshalb wurde anschließend ein zusätzliches Linkage mit Blocking auf Geschlecht, Geburtsjahr und Geburtsmonat durchgeführt.

Bei der Berechnung des Ähnlichkeitsmaßes wurde ein probabilistisches Linkage angewandt, das rechnerisch berücksichtigt, dass die Wahrscheinlichkeit einer Übereinstimmung von Merkmalen von deren Varianz in der Grundgesamtheit abhängig ist. Hierbei werden so genannte m- und u-Wahrscheinlichkeiten 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 eine Übereinstimmung beim Nachnamen und wird deshalb niedriger gewichtet. Die variablenspezifischen m- und u-Gewichte, die dem probabilistischen Linkage zugrunde lagen, wurden im Rahmen vergangener Projekte des GRLC ermittelt. Es wurden alle zur Verfügung stehenden Merkmale zwischen den beiden Datenquellen verglichen und ein additives Ähnlichkeitsmaß berechnet. Je höher der Wert dieses Index, desto größer die Gesamtähnlichkeit der beiden Zeilen und desto wahrscheinlicher liegt eine tatsächliche Übereinstimmung vor.

Von allen verglichenen Zeilenpaaren wurden nur jene oberhalb eines festgelegten Schwellenwerts des Ähnlichkeitsmaßes als potentielle Matches betrachtet. Von diesen potenziellen Übereinstimmungen wurde pro befragter Person jeweils die Entsprechung in den IEB-Daten mit der höchsten Gesamtähnlichkeit übernommen. Mit anderen Worten wurde also der beste Match für jede Person ausgewählt, sofern dieser eine hinreichend hohe Qualität aufwies. Der Schwellenwert wurde dabei so gewählt, dass zum einen die Toleranz für Abweichungen hinreichend groß ist, damit tatsächliche Übereinstimmungen auch als Match gewertet werden, zum anderen aber nicht zu niedrig, da damit die Wahrscheinlichkeit von Fehlzusordnungen zunehmen würde.

Für die durch das fehlertolerante Linkage zugeordneten Personen bestehen kleinere Abweichungen bei der Schreibweise der Felder Vor- und Nachnamen, Straße oder Ort. Je kleiner die Abweichung, desto höher das berechnete Ähnlichkeitsmaß. Indem nur Beobachtungen mit einem Übereinstimmungsmaß oberhalb eines festgelegten Schwellenwerts als potentielle Matches betrachtet wurden, war die Wahrscheinlichkeit sehr gering, im Rahmen des fehlertoleranten Linkage unterschiedliche Personen als Übereinstimmung zu identifizieren. Geschlecht und Geburtsdatum der Personen stimmten in 95% der Fälle überein. Nur sehr selten wies entweder eine Ziffer beim Geburtsdatum ab oder es gab divergierende Angaben zum Geschlecht der Person. Eine manuelle Sichtung ergab jedoch, dass es sich in diesen Fällen um plausible Matches handelte, bei denen im Ursprungsmaterial Fehlangaben vorlagen.

Für die verbleibenden Personen, die durch keines der o. g. Verfahren gematcht werden konnten, wurde ein manuelles Linkage durchgeführt. Hierzu wurden die Zeilen unterhalb der Qualitätsgrenze manuell gesichtet und plausible Matches markiert.

Zur Überprüfung der Zuordnungen wurden verschiedene Linkage-Schritte durchgeführt und die resultierenden Personenverknüpfungen verglichen, wobei eine hohe Reliabilität der Zuordnungen festgestellt wurde. Zusätzlich werden stichprobenweise manuelle Sichtungen vorgenommen, um die Plausibilität der Zuordnungen zu kontrollieren. Hierbei wurden besonders diejenigen Verknüpfungen kontrolliert, bei denen einige Felder nicht übereinstimmen.

## 4.6 Ergebnis der Datenverknüpfung

Insgesamt 12.374 Personen erteilten bis einschließlich Welle 12 ihre Zustimmung zur Verknüpfung (69,09% aller Befragten). Davon konnten jedoch nur 11.340 Personen für die Datenverknüpfung verwendet werden, da Personen, die ihre Panelbereitschaft zwischenzeitlich zurückgezogen haben, für das Linkage aus rechtlichen Gründen nicht zur Verfügung standen, obgleich sie der Panelbereitschaft zuvor zugestimmt hatten. Personen, die vor Welle 5<sup>14</sup> ihre Panelbereitschaft zurückgezogen haben und dadurch nie nach ihrer Zustimmung zur Verknüpfung befragt wurden, werden als „der Verknüpfung nicht zugestimmt“ behandelt.

Für 10.468 Personen konnten die Erhebungsdaten einem Konto in den administrativen Daten zugeordnet werden. Dies entspricht einer sehr hohen Verknüpfungsquote von 92,31% der Personen, die für eine Zuspiegelung der administrativen Daten zur Verfügung standen. Es verblieben lediglich 872 Personen (7,69%), die trotz Verknüpfungserlaubnis nicht in den administrativen Datenbeständen<sup>15</sup> gefunden wurden. In Tabelle 8 werden die Erfolgsquoten der einzelnen Verknüpfungsschritte zusammengefasst.

Generierte technische Merkmale im Datensatz (siehe Kapitel 6.8) kennzeichnen für jede befragte Person, mit welcher Verknüpfungsmethode und mit welchem Maß an Sicherheit eine erfolgreiche Verknüpfung zustande gekommen ist. Damit kann jede Nutzerin und jeder Nutzer nach eigenem Ermessen entscheiden, welche der Fälle in den eigenen Analysen Verwendung finden sollen.

Tabelle 8 Ergebnisse des Linkage-Prozess

	N	Anteil an Zustimmenden	Anteil an Befragten
Befragte	17.910		100,00%
Befragte mit Verknüpfungserlaubnis	12.374		69,09%
verfügbare Befragte	11.340	100,00%	63,32%
Matches mit IEB-Sätzen	10.468	92,31%	58,45%
darunter:			
Exaktes Matching	10.219	90,11%	57,06%
Probabilistisches Matching	222	1,96%	1,24%
Manuelles Matching	27	0,24%	0,15%

<sup>14</sup> Die Zustimmung zur Verknüpfung wurde erstmals in Welle 5 erhoben.

<sup>15</sup> Für 673 Personen, die gelinkt wurden, konnten keine administrativen Daten gefunden werden. Aus diesem Grund sind sie in der folgenden Tabelle nicht als Matches berücksichtigt

# 5 Datenqualität und Probleme

## 5.1 NEPS-SC5-Befragungsdaten (NEPS-SC5)

Eine präzise Beschreibung der Daten, deren Struktur sowie Informationen zur Stichprobe und zur Befragung allgemein finden sich im Data Manual der Startkohorte 5.<sup>16</sup> Weiterhin sind dort generierte Variablen und Details zur Gewichtung aufgeführt.<sup>17</sup>

## 5.2 Integrierte Erwerbsbiografien (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 kennen (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 (s. Tabelle 9). Die Auflistung in Tabelle 9 erhebt keinen Anspruch auf Vollständigkeit.

Tabelle 9 Biographische Lücken und Identifikationsmöglichkeiten

Biografische Lücke	Information über Lücke, ggf. identifizierbar anhand der Ausprägung des Merkmals „grund“ im Vorgängersatz der Quelle
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

<sup>16</sup>Siehe <https://www.neps-data.de>.

<sup>17</sup> Im Rahmen des Verknüpfungsprozesses wurde eine Dublette in den Daten der SC5 identifiziert. Es liegen zwei unterschiedliche IDs vor (ID\_t=7011365 und ID\_t=7016685), bei denen es sich mit sehr hoher Wahrscheinlichkeit um dieselbe Person handelt. Da seitens des NEPS eine Auflösung der Dublette erst mit der nächsten Version der Daten möglich ist, sind in den Daten von NEPS-SC5-ADIAB beide IDs enthalten. Für diese IDs liegen folglich identische Informationen in den administrativen Daten vor.

Personen, die länger als 6 Wochen krank / arbeitsunfähig sind (Krankheit während Arbeitslosigkeit ist aber unter bestimmten Bedingungen in der Quelle ASU abgebildet, s. Abschnitt 3.1.6)	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 SGB 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 Übermittlung von Informationen zur Ausbildung, ausgeübten Tätigkeit und Stellung im Beruf erfolgt durch Meldungen der Arbeitgeber nach DEÜV (siehe Kapitel 2.3) über einen sogenannten Tätigkeitsschlüssel. Für Zeiträume mit einem Enddatum nach dem 30.11.2011 erfolgte der Wechsel zum neuen Tätigkeitsschlüssel 2010 (siehe Bertat et al., 2013, für mehr Details). Dieser Wechsel wurde von den Spitzenorganisationen der Sozialversicherungsträger beschlossen, da der Tätigkeitsschlüssel 2003 eine Reihe von Sachverhalten nicht mehr zeitgemäß und nicht mehr realitätsnah erfassen konnte. Da die Meldungen der Arbeitgeber nach DEÜV in die BeH einfließen, hat die Umstellung des Tätigkeitsschlüssels ausschließlich Auswirkungen für Beobachtungen aus dieser Quelle. Die Messung folgender bisher enthaltener An-

gaben ist von dieser Umstellung betroffen: Arbeitszeit, Beruf, Stellung im Beruf sowie schulisches und berufliches Qualifikationsniveau. Darüber hinaus gibt es seit der Umstellung auch Angaben zur Befristung eines Beschäftigungsverhältnisses sowie dazu, ob eine Person als Verleihpersonal in einem Zeitarbeitsunternehmen angestellt ist. Die wichtigste Auswirkung ist der Umstieg auf eine neue Berufsklassifikation. Statt der bisherigen Klassifikation der Berufe 1988 (KldB 1988) wird mit dem neuen Tätigkeitsschlüssel die differenziertere KldB 2010 gemeldet.<sup>18</sup>

- Anders als in der SIAB 7514 ist das Merkmal „Stellung im Beruf und Arbeitszeit“ (stib) in der SIAB 7517 wieder enthalten. Die Variable ist allerdings nur für Meldungen gefüllt, die vor Einführung des neuen Tätigkeitsschlüssels abgegeben wurden, und das FDZ nimmt keine Extrapolation oder Imputation für spätere Meldungen vor. Die Kategorien der Variable stib, die über den gesamten Beobachtungszeitraum konsistent beobachtet werden können (Erwerbsstatus als Auszubildender, Unterscheidung von Teil- und Vollzeit), sind in den Variablen erwstat und teilzeit nachzuvollziehen.
- Die Einführung des neuen Tätigkeitsschlüssels hat zu einer Reihe von Problemen geführt. So kam es während des Übergangszeitraums, der Arbeitgebern im Meldeverfahren zur Sozialversicherung eingeräumt wurde,<sup>19</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 7517 verwendet. Für die Lücken in den anderen Merkmalen werden keine Imputationen vorgenommen.
- 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 ist 2014 das letzte Jahr mit einem Füllgrad der BeH-Meldungen von 100%.<sup>20</sup> Für die Jahre 2015 und 2016 wurden 18-Monats-Dateien verwendet,

---

<sup>18</sup> Siehe Paulus/Matthes, 2013, für Details zur Klassifikation der Berufe 2010.

<sup>19</sup> Die im Rahmen des Meldeverfahrens eingesetzten Prüfprogramme ließen bis einschließlich Mai 2012 leere Angaben beim Tätigkeitsschlüssel 2010 zu.

<sup>20</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.

für 2017 eine 6-Monats-Datei. Es ist daher davon auszugehen, dass Beschäftigungsmeldungen für die Jahre 2015 bis 2016 leicht und für 2017 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.

## 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 verfügbar. Daher ist für das vorliegende Datenprodukt auch davon auszugehen, dass die Informationen zum Leistungsbezug im genannten Zeitraum nur zum Teil enthalten sind.

## 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 01.01.2011 kommt es zu Untererfassungen im Gebiet der Arbeitsagenturen Dessau-Roßlau, Halberstadt, Halle und Sangerhausen.
  - Doppelmeldungen aufgrund der Gebietsreformen 2009/2011 und der Trägerformwechsel zum 01.01.2012 und zum 01.01.2013 werden in der IEB bereits soweit möglich bereinigt. Dennoch können Doppelmeldungen auftreten.
- Bei den 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.

- Aufgrund der Meldelogik sind Informationen aus dem Meldeverfahren XSozial nur monatsgenau zuspelbar.
- In der offiziellen Leistungsstatistik der BA werden Meldelücken durch ein statistisches Schätzverfahren auf aggregierter regionaler Ebene ergänzt. In der LHG werden keine ergänzenden Datensätze bereitgestellt. Ein Vergleich der IEB mit der Leistungsstatistik ist daher z.T. nur eingeschränkt möglich.

## 5.6 Arbeitssuchendenhistorik (ASU/XASU)

### 5.6.1 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.
- Mit der Einführung des SGB II zum 01.01.2005 werden arbeitssuchende Personen nicht mehr vollständig mit BA-Verfahren erfasst. Die Quelle ASU berücksichtigt ab diesem Zeitpunkt nur noch Personen, die von der BA im Rechtskreis SGB III (Arbeitsförderung), oder von ARGEn, gE bzw. gT im Rechtskreis SGB II (Grundsicherung) betreut werden.
- 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 Arbeitssuchendendaten 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. Es ist von einer Untererfassung von Eingliederungsvereinbarungen auszugehen. Wichtige Einschränkungen des Analysepotentials werden in der jeweiligen Variablenbeschreibung angesprochen.
- Das bis Juni 2006 genutzte Verfahren coArb unterstützte ausschließlich die Vermittlung von Arbeitslosen und Arbeitssuchenden. 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 eingeschränkt für die Arbeitsvermittlung verfügbar sind. 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.
- Der Arbeitsuche-Status wird nachträglich kaum korrigiert, weshalb die Fallzahlen mit der BA-Statistik bis 2005 weitgehend übereinstimmen. Seit Frühjahr 2011 werden die Arbeitsuchendendaten aus BA-Verfahren und XSozial in der integrierten Arbeitslosenstatistik konsolidiert verarbeitet. Hier kann es zu größeren Abweichungen kommen. Am aktuellen Rand sind die Bestandsdaten jedoch identisch.

### 5.6.2 XASU

- Im Gegensatz zu den Arbeitsuchendendaten aus den ARGEn und den getrennten Trägerschaften zeigen sich nach 01.01.2005 für die zugelassenen kommunalen Träger zwischenzeitlich systematische Untererfassungen. Deshalb sollten Daten aus der Quelle XASU erst ab 2007 ausgewertet werden.
- Auch in den Jahren ab 2007 sind analog zu den Leistungsdaten (LHG) immer wieder einzelne Monate von Lieferausfällen betroffen.
- Punktuell kann es auch zu Übererfassungen kommen. Hierfür kann eine Vielzahl technischer Probleme verantwortlich sein, u.a. systematisch nicht erfolgte Abmeldungen, falsche Entstornierungen und parallele Arbeitsuchendenmeldungen in mehreren Kreisen.
- 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 unplausibel 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.
- Aufgrund der Meldelogik sind Informationen aus dem Meldeverfahren XSozial nur monatsgenau zuspielbar.
- Durch Unterschiede in Konsolidierungsregeln, Zeitbezügen und regionalen Zuordnungen kann es zu Unterschieden zu den veröffentlichten BA-Statistiken kommen.

## 5.7 Maßnahmeteilnahmehistoriken (MTH/XMTH)

### 5.7.1 MTH

- Die MTH sind für Maßnahmen mit einem Eintrittsdatum vor dem 01.01.2000 unvollständig.
- Zum 01.05.2005 kommt es zu einem Bruch, da mit Einführung des SGB II Maßnahmeteilnehmerinnen und -teilnehmer auf verschiedene Träger aufgeteilt wurden (siehe Kapitel 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 NEPS-SC5-ADIAB enthalten.
- Durch die Neuorganisation von SGB-II-Trägern 2011-2014 kann es zur Aufspaltung der Dokumentation von Maßnahmeteilnahmen in der MTH und XMTH kommen, wenn es zu einer Umstellung des Meldeverfahrens kommt.<sup>21</sup> Dies kann sich in einem Split des Maßnahmespells, aber auch in Dopplungen niederschlagen.
- 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.

### 5.7.2 XMTH

- In den Jahren 2005 bis 2007 sind die Meldungen von Maßnahmeteilnahmen lückenhaft. Die Höhe der Untererfassung ist wegen fehlender bzw. ungenauer Vergleichszahlen unklar.
- Ca. 13 % aller Grundsicherungsträger melden von 2005 bis Anfang 2017 fast keine kommunalen Eingliederungsleistungen (vormals flankierende Leistungen), die u.a. in der zusammengefassten Kategorie „sonstige Förderung“ enthalten sind. Viele weitere Jobcenter melden nur zeitweise und / oder nur ausgewählte Arten. Das führt zu einer Untererfassung der „sonstigen Förderung“.
- Übererfassung von Teilnahmen: Die Gesamtbestände wirken ab Ende 2008 stabil. Bei der Mehrzahl der Träger sind aber auch danach zeitweise auffällige Lieferprofile zu registrieren, die Zweifel an der Qualität wecken. Erst die Einführung der XSozial-Förder-ID im Jahr 2009 und die mittlerweile mehrjährige Erfahrung aller Beteiligten stabilisieren das Meldegeschehen und die daran anschließende Datenverarbeitung. Auch die ab 2012 neu hinzukommenden zkT melden weit überwiegend unauffällig, so dass Umfang und Dauer mutmaßlicher Übererfassung merklich zurückgehen.

---

<sup>21</sup> Weitere Informationen zur Gebietsstruktur der SGB-II-Träger und relevanten Änderungen finden sich unter: <https://statistik.arbeitsagentur.de>.

- Durch die Neuorganisation von SGB-II-Trägern 2011-2014 kann es zur Aufspaltung der Dokumentation von Maßnahmeteilnahmen in der MTH und XMTH kommen, wenn es zu einer Umstellung des Meldeverfahrens kommt. Dies kann sich in einem Split des Maßnahmespells, aber auch in Dopplungen niederschlagen (siehe Abschnitt 3.1.8).
- Die über monatliche Stichtagsauszählungen in der XMTH erhaltenen Zahlen bzw. Personen unterscheiden sich in mehrfacher Hinsicht von den veröffentlichten BA-Statistiken.

## 6 Variablenbeschreibung

Auszählungen sowie Übersichten über die einzelnen Ausprägungen und Labels der Merkmale finden sich in den Arbeitshilfen zu NEPS-SC5-ADIAB unter <http://fdz.iab.de>.

### 6.1 Identifikatoren

#### 6.1.1 Personen-ID (ID\_t)

Variablenlabel	ID-Target
Variablenname	ID_t
Kategorie	Identifikatoren
Herkunft	Generiert
Datentyp	Numerisch
Detailbeschreibung	Die Personen-ID zeigt an, welche Sätze zu einer Person gehören. Sie zeigt auch an, welche Personen in Erhebungs- und administrativen Daten korrespondieren, d.h. die ID_t ist sowohl in den Erhebungsdaten als auch in den administrativen Daten enthalten. Diese Identifikationsvariable ist systemfrei, weshalb sie keine Rückschlüsse auf Eigenschaften der Person oder Original-Identifikatoren ermöglicht.

#### 6.1.2 Betriebs-ID (betnr)

Variablenlabel	Systemfreie Betriebsnummer
Variablenname	betnr
Kategorie	Identifikatoren
Herkunft	BEH
Datentyp	Numerisch
Detailbeschreibung	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) sowie direkt bei dem Betriebsnummern-Service der BA unter <a href="https://www.arbeitsagentur.de/betriebsnummern-service/alles-wichtige">https://www.arbeitsagentur.de/betriebsnummern-service/alles-wichtige</a> ). Die Betriebsnummer kann zusammen mit der Jahresangabe benutzt werden, um Personen- und Betriebsinformationen zusammenzuspielen. Für die Betriebsnummer sollte Folgendes grundsätzlich beachtet werden: 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.

	<p>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.</p> <p>Hat das Unternehmen <b>mehrere</b> Niederlassungen in <b>mehreren</b> Gemeinden, so ist jede dieser Niederlassungen ein Betrieb und erhält je eine Betriebsnummer.</p> <p>In diesem Zusammenhang sind auch folgende Definitionen im Sinne der Betriebsnummernvergabe im Rahmen des Meldeverfahrens zur Sozialversicherung zu beachten:</p> <p>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 vergeben wird.</p> <p>Eine <b>Arbeitsstätte</b> ist eine Einheit, in der Beschäftigte tätig sind und für die nach o.a. Grundsätzen keine Betriebsnummer zu vergeben ist.</p> <p>Ein <b>Unternehmen</b> als Begriff fasst Betriebs- und Arbeitsstätten desselben Arbeitgebers zusammen.</p> <p>Ein <b>Arbeitgeber</b> ist jede natürliche oder juristische Person, für die mindestens ein sozialversicherungspflichtig bzw. geringfügig Beschäftigter tätig ist.</p> <p>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.</p>
Qualitätshinweis	<p>Die Betriebsnummer fehlt in nur sehr wenigen BeH-Meldungen. Bei diesen Meldungen handelt es sich in den meisten Fällen um Meldungen der Personengruppe „205“ (Entgeltmeldungen für unständig Beschäftigte). Da Betriebsmerkmale (Arbeitsort, Wirtschaftszweig, Betriebsgröße usw.) über die Betriebsnummer zugespielt werden, fehlen sie in diesen Meldungen.</p>

## 6.2 Generierte technische Merkmale

### 6.2.1 Satzzähler pro Konto (spell)

Variablenlabel	Satzzähler pro Konto
Variablenname	spell
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Detailbeschreibung	<p>Der Satzzähler pro Konto nummeriert aufsteigend die Sätze einer Person. Das Merkmal wird beim Episodensplitting erzeugt und bezieht sich auf die gesplitteten Sätze. Mit dem Merkmal „Satzzähler pro 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. Innerhalb der Beschäftigungsmeldungen werden sozialversicherungspflichtige vor geringfügige Beschäftigungsmeldungen und höhere vor niedrigere Tagesentgelte sortiert. Eine Ausnahme bilden Einmalzahlungen, die nach hinten sortiert werden.</p>

### 6.2.2 Quelle des Satzes (quelle)

Variablenlabel	Quelle des Satzes
Variablenname	quelle
Kategorie	Generierte technische Merkmale

Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal gibt die Datenquelle an (s. Kapitel 2).

### 6.2.3 Jahr (jahr)

Variablenlabel	Jahr
Variablenname	jahr
Kategorie	Generierte technische Merkmale
Herkunft	BeH
Datentyp	Numerisch
Detailbeschreibung	<p>Dieses Merkmal ist nur in der Betriebsdatei enthalten. Es gibt das Gültigkeitsjahr der Betriebsinformationen zum Stichtag 30.06. an.</p> <p>Mit Hilfe dieses Merkmals und der Betriebsnummer können Personen- und Betriebsdatei verknüpft werden. Siehe Kasten 1 im Kapitel 2.3 für einen Beispielcode mit dem „merge“-Befehl in Stata 14.</p>

## 6.3 Gültigkeitszeitraum

### 6.3.1 Beginndatum Originalbeobachtung (begorig)

Variablenlabel	Beginndatum Originalbeobachtung
Variablenname	begorig
Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH , XMTH
Datentyp	Datum
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 Kapitel 3.2). Da die Meldelogik unter bestimmten Umständen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben in diesen Fällen durch das im Abschnitt 3.1.9 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 Kapitel 2.5 und 2.6). begorig gibt das Beginndatum des neuen Zeitraums an.</p> <p><b>3) MTH, XMTH</b> In der MTH und XMTH entstehen neue Datensätze, wenn ein Trägerwechsel während der Teilnahme an einer arbeitsmarktpolitischen Maßnahme stattfindet.</p>

### 6.3.2 Enddatum Originalbeobachtung (endorig)

Variablenlabel	Enddatum Originalbeobachtung
Variablenname	endorig

Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Datum
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 Kapitel 3.2). Da die Meldelogik unter bestimmten Umständen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben in diesen Fällen durch das im Abschnitt 3.1.9 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 Kapitel 3.1.5 und 3.1.6). endorig gibt das Enddatum des neuen Zeitraums an.</p> <p><b>3) MTH, XMTH</b> In der MTH und XMTH entstehen neue Datensätze, wenn ein Trägerwechsel während der Teilnahme an einer arbeitsmarktpolitischen Maßnahme stattfindet.</p>

### 6.3.3 Beginndatum der Episode (begepi)

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

### 6.3.4 Enddatum der Episode (endepe)

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

## 6.4 Informationen zur Person

### 6.4.1 Geschlecht (frau)

Variablenlabel	Geschlecht
Variablenname	frau

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

#### 6.4.2 Geburtsjahr (gebjahr)

Variablenlabel	Geburtsjahr
Variablenname	gebjahr
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
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.

#### 6.4.3 Geburtsmonat (gebmon)

Variablenlabel	Geburtsmonat
Variablenname	gebmon
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	Der Geburtsmonat ist innerhalb eines Personenkontos konstant. Um aus den Angaben zum Geburtsjahr (gebjahr) und -monat (gebmon) eine Variable im Datumsformat JJJJmM (z.B. 1984m6) zu generieren, kann folgender Befehl verwendet werden: <pre>gen int gebdat = ym(gebjahr, gebmon) format gebdat %tm</pre>
Qualitätshinweis	In den Originaldaten kann es vorkommen, dass das Geburtsdatum zwischen den Quellen wechselt. Dies wird bei der Datenaufbereitung bereinigt. Hierbei kommt Angaben aus dem Leistungsbezug die höchste Priorität zu.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen bereitgestellt. Ansonsten wird lediglich das Geburtsjahr (gebjahr) bereitgestellt.

#### 6.4.4 Staatsangehörigkeit (nation)

Variablenlabel	Staatsangehörigkeit
Variablenname	nation
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält als Ausprägungen die Staatenschlüssel des Statistischen Bundesamtes (Statistisches Bundesamt, 2020).
Qualitätshinweis	Das Merkmal ist in der LeH erst ab 1983 gut gefüllt.

Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird als Staatsangehörigkeitsmerkmal lediglich die Staatsangehörigkeit vergrößert (nation_gr) bereitgestellt.
----------------	---

#### 6.4.5 Staatsangehörigkeit vergrößert (nation\_gr)

Variablenlabel	Staatsangehörigkeit vergrößert
Variablenname	nation_gr
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält als Ausprägungen die zu Gruppen zusammengefassten Staatenschlüssel des Statistischen Bundesamtes (Statistisches Bundesamt, 2020).
Qualitätshinweis	Das Merkmal ist in der LeH erst ab 1983 gut gefüllt.

#### 6.4.6 Familienstand (famst)

Variablenlabel	Familienstand
Variablenname	famst
Kategorie	Persönliche Merkmale
Herkunft	LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	<p>Das Merkmal beschreibt den Familienstand.</p> <p><b>1) LeH</b> In der Quelle LeH umfasst das Merkmal nur zwei Ausprägungen (0 - nicht verheiratet, 1 - verheiratet).</p> <p><b>2) LHG, ASU, XASU, MTH, XMTH</b> In den Quellen LHG/ASU/XASU/MTH/XMTH wird zwischen sechs Ausprägungen (Werte 11-16) unterschieden. Ein Abgleich der Informationen aus den Quellen wurde nicht vorgenommen.</p>
Qualitätshinweis	<p><b>1) LeH</b> Die Qualität des Merkmals aus der LeH wird bis 2005 als schlecht eingestuft, da nicht ganz geklärt ist, wie der Familienstand gepflegt wurde. Auch ab Beginnjahr 2014 ist die Qualität fraglich. So sinkt der Anteil der Verheirateten z.B. zwischen 2013 und 2017 von 37% auf 14%.</p> <p><b>2) LHG, XASU, XMTH</b> Aufgrund abweichender Meldestandards in XSozial sind die Angaben aus den zkT bis Dezember 2009 als weniger valide einzustufen.</p>

#### 6.4.7 Kinderzahl (kind)

Variablenlabel	Kinderzahl
Variablenname	kind
Kategorie	Persönliche Merkmale
Herkunft	LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	<p>Je nach Datenquelle hat das Merkmal eine unterschiedliche Bedeutung.</p> <p><b>1) LeH</b></p>

	<p>In der LeH gibt die Variable die Anzahl der Kinder unter 16 Jahren an, die zum Zeitpunkt der Antragstellung im Haushalt leben. Es wird nur unterschieden, ob die Anzahl an Kindern gleich Null oder mindestens eins ist. Folgende Ausprägungen kommen daher vor:</p> <p>0      Keine Kinder 100    Ein oder mehrere Kinder</p> <p><b>2) ASU, MTH</b> Hier entspricht die Ausprägung des Merkmals der Anzahl an Kindern unter 15 Jahren im Haushalt.</p> <p><b>3) LHG, XASU, XMTH</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>
Qualitätshinweis	<p><b>1) LeH</b> Das Merkmal wird bei Änderungen der Leistungsart oder neuen Bewilligungen nicht aktualisiert, sondern erst, wenn ein erneuter Leistungsfall auftritt. Dadurch ist die Datenqualität stark eingeschränkt.</p> <p><b>2) ASU, MTH</b> Bis zum 30.06.2006 konnten nur bis zu neun Kinder erfasst werden. Der Wert Null kommt nicht vor. Vor dem 30.06.2006 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 30.06.2006 ist das Merkmal nur gefüllt, wenn Kinder vorhanden sind.</p>

#### 6.4.8 Ausbildung (ausbildung)

Variablenlabel	Ausbildung
Variablenname	ausbildung
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
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> <ol 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> </ol> <p>Bei Meldungen nach dem neuen Tätigkeitsschlüssel (siehe Kapitel 2.3) 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><b>2) ASU, MTH</b> Für diese Sätze ist die letzte abgeschlossene Berufsausbildung enthalten. Folgende Ausprägungen kommen bei Spells mit Beginndatum bis zum 30.06.2006 vor:</p> <ol style="list-style-type: none"> <li>1    ohne abgeschlossene Berufsausbildung</li> <li>2    betriebliche/außerbetriebliche Ausbildung</li> <li>3    Berufsfachschule</li> </ol>

- 4 Fachschule
- 5 Fachhochschule
- 6 Universität

Im Jahr 2006 wurde das EDV-Verfahren, aus dem die Arbeituchendendaten stammen, von coArb auf VerBIS umgestellt. Viele Merkmale, wie auch die Ausbildung, wurden in beiden Systemen in unterschiedlicher Differenziertheit abgefragt. Das bedeutet, dass in den Quellen ASU und MTH bei Spells mit Beginndatum ab dem 01.07.2006 folgende differenzierteren Aufprägungen vorliegen:

- 7 in Deutschland nicht anerkannte Berufsausbildung
- 8 in Deutschland nicht anerkannter Hochschulabschluss
- 9 betriebliche/schulische Ausbildung
- 10 sonstige Examen
- 11 FH ohne nähere Spezifikation
- 12 Hochschulabschluss ohne nähere Spezifikation
- 13 Promotion
- 14 Bachelor (BA)
- 15 Bachelor (FH)
- 16 Bachelor (Uni)
- 17 Master (FH)
- 18 Master (Uni)
- 19 Diplom (BA)
- 20 Diplom (FH)
- 21 Diplom (Uni)
- 23 Grundständige Studienfächer/-gänge
- 24 Weiterführende Studienfächer/-gänge
- 25 Sonstige (Aufstiegs-)Weiterbildungen für Hochschulabsolventen

### 3) XASU, XMTH

Für diese Sätze ist die letzte abgeschlossene Berufsausbildung enthalten. Folgende Ausprägungen kommen vor:

- 1 ohne abgeschlossene Berufsausbildung
- 2 betriebliche/außerbetriebliche Ausbildung
- 3 Berufsfachschule
- 4 Fachschule
- 5 Fachhochschule
- 6 Universität
- 7 in Deutschland nicht anerkannte Berufsausbildung
- 8 in Deutschland nicht anerkannter Hochschulabschluss

Für Auswertungen über einen längeren Zeitraum können die Ausprägungen anhand folgender Tabelle aggregiert werden:

Ausbildung	ausbildung_agg	Variablenbezeichnung
1, 22	1	ohne (anerkannte) abgeschlossene Berufsausbildung
7	2	in Deutschland nicht anerkannte Berufsausbildung
8	3	in Deutschland nicht anerkannter Hochschulabschluss
2, 3, 4, 9	4	betriebliche / schulische Ausbildung
5, 11, 14, 15, 17, 19, 20	5	Fachhochschule
6, 12, 13, 16, 18, 21	6	Universität
10	6	sonstige Examen
23	7	Grundständige Studienfächer/-gänge

	24	7	Weiterführende Studienfächer/-gänge
	25	7	Sonstige (Aufstiegs-)Weiterbildungen für Hochschulabsolventen
Qualitätshinweis	<p><b>1) BeH</b></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>Der Anteil der fehlenden Werte steigt in der BeH im Zeitverlauf fast kontinuierlich an. Durch die Einführung des neuen Tätigkeitsschlüssels gab es in 2011 einen stark erhöhten Anteil von rund 51%. Seit 2014 pendelt sich der Anteil der fehlenden Werte in der BeH bei rund 40% ein. Fehlende Werte treten besonders häufig bei den folgenden Personengruppen auf: geringfügig Beschäftigte, 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>Für das Merkmal Ausbildung imputiert (ausbildung_imp, siehe Abschnitt 6.4.9) wurde eine Methode angewendet, um fehlende Werte oder inkonsistente zeitliche Verläufe der Bildungsangaben zu bereinigen. Diese Variable ist jedoch nur in der Quelle BeH gefüllt und hat andere Ausprägungen als die Variable ausbildung.</p> <p><b>2) ASU, MTH</b></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. In der Quelle MTH sind auch nach dem 01.07.2006 zum Teil noch die älteren Ausprägungen zu finden.</p> <p><b>3) XASU, XMTH</b></p> <p>Der Füllgrad in den Quellen XASU und XMTH ist generell niedrig.</p>		

#### 6.4.9 Ausbildung (imputiert) (ausbildung\_imp)

Variablenlabel	Ausbildung (imputiert)
Variablenname	ausbildung_imp
Kategorie	Persönliche Merkmale
Herkunft	BeH
Datentyp	Numerisch

Detailbeschreibung	<p>Das Merkmal Ausbildung (imputiert) ist ein Zusatzmerkmal zum Merkmal Ausbildung (ausbildung) und enthält für BeH-Meldungen ergänzte und harmonisierte Angaben zur Berufsausbildung von Beschäftigten. Das Merkmal bietet damit einen Lösungsansatz für die in Abschnitt 6.4.8 beschriebenen Probleme der Variable Ausbildung (ausbildung) bei BeH-Meldungen. Das Imputationsverfahren wird in Thomsen et al. (2018) beschrieben, basierend auf der Arbeit von Fitzenberger et al. (2006).</p> <p>Da das Merkmal nur Bildungsangaben der BeH-Meldungen nutzt, und weil die Bildungskategorien des alten und des neuen Tätigkeitsschlüssels für das Merkmal Ausbildung (ausbildung) harmonisiert werden mussten, bietet das Merkmal Ausbildung (imputiert) (ausbildung_imp) andere Kategorien als das Merkmal Ausbildung (ausbildung).</p>
--------------------	---

#### 6.4.10 Schulausbildung (schule)

Variablenlabel	Schulausbildung
Variablenname	schule
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	<p>Das Merkmal enthält den Schulabschluss. Je nach Quelle sind unterschiedliche Ausprägungen möglich.</p> <p><b>1) BeH</b></p> <p>Mit der Umstellung des Tätigkeitsschlüssels ändern sich die möglichen Ausprägungen (siehe Kapitel 2.3). Davor gelten:</p> <ul style="list-style-type: none"> <li>5 Volks- / Hauptschule o./m. Abschluss, Mittlere Reife oder gleichwertige Schulbildung</li> <li>8 Fachhochschulreife oder Abitur/Hochschulreife</li> <li>9 Abitur / Hochschulreife</li> </ul> <p>Danach gelten:</p> <ul style="list-style-type: none"> <li>1 Kein Hauptschulabschluss</li> <li>4 Hauptschulabschluss / Volksschulabschluss</li> <li>6 Mittlere Reife</li> <li>8 Fachhochschulreife oder Abitur / Hochschulreife</li> </ul> <p><b>2) ASU, XASU, MTH, XMTH</b></p> <p>Für diese Quellen sind die folgenden Ausprägungen möglich:</p> <ul style="list-style-type: none"> <li>1 Kein Hauptschulabschluss</li> <li>4 Hauptschulabschluss / Volksschulabschluss</li> <li>6 Mittlere Reife</li> <li>7 Fachhochschulreife</li> <li>9 Abitur / Hochschulreife</li> </ul> <p>Sie sind zu Beginn der Arbeitsuche / Maßnahmeteilnahme gültig. In der Quelle XASU kann das Merkmal bei Ausbildungsplatzsuchenden auch den angestrebten Abschluss enthalten.</p>
Qualitätshinweis	<p>Der Füllgrad nimmt in der BeH über die Jahre kontinuierlich ab und liegt in jüngeren Jahren unter 2/3. In den Quellen XASU und XMTH nimmt er kontinuierlich zu und liegt seit 2013 (XASU) bzw. seit 2012 (XMTH) über 2/3. Das Merkmal kann in den Quellen XASU und XMTH für Ausbildungsplatzsuchende auch den angestrebten Schulabschluss enthalten. In ASU und MTH ist der Füllgrad insgesamt gut.</p>

## 6.5 Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

### 6.5.1 Tagesentgelt / täglicher Leistungssatz (tentgelt)

Variablenlabel	Tagesentgelt / täglicher Leistungssatz
Variablenname	tentgelt
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH
Datentyp	Numerisch
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 01.04.1999 werden auch Entgelte unterhalb der Geringfügigkeitsgrenze erfasst. Die Beitragsbemessungsgrenze als obere Kappungsgrenze 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="https://fdz.iab.de">https://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>Ab 2013 nimmt die Anzahl der Meldungen mit Abmeldegrund 54 (Meldung eines einmalig gezahlten Entgelts) sprunghaft zu (siehe Abschnitt 6.5.12). Es ist davon auszugehen, dass Sonderzahlungen, die vor 2013 mit den Jahresmeldungen angegeben wurden, nun extra gemeldet werden. Daher empfiehlt es sich bei Analysen der Entgelte im Zeitverlauf, die Sonderzahlungen zu den entsprechenden Entgelten von zeitgleichen Beschäftigungsepisoden im gleichen Betrieb dazu zu addieren.</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 um-</p>

	<p>gerechnet in Euro. Es ist zu berücksichtigen, dass der tägliche Leistungssatz bei Datensätzen mit einem Original-Enddatum, das vor dem 01.01.1998 liegt, für Werktage (also ohne Sonn- und Feiertage, jedoch inklusive Samstage) und bei Datensätzen mit einem Original-Enddatum ab 01.01.1998 für Kalendertage gilt.</p> <p>Die Angabe von 0 Euro ist seit 01.01.2005 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>
--	---

### 6.5.2 Beruf – ausgeübte / letzte Tätigkeit (KldB 1988) (beruf)

Variablenlabel	Beruf – ausgeübte / letzte Tätigkeit (KldB 1988)
Variablenname	beruf
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, ASU, XASU, MTH
Datentyp	Numerisch
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 eines 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 (siehe Kapitel 2.3). 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><b>1) BeH</b> In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf.</p> <p><b>2) ASU, MTH</b> Der Füllgrad fällt in den Jahren 2014 bis 2017 deutlich ab.</p> <p><b>3) XASU</b> Das Merkmal Beruf ist nahezu für den gesamten zur Verfügung stehenden Zeitraum nicht gefüllt.</p>

### 6.5.3 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, XMTH
Datentyp	Numerisch

Detailbeschreibung	<p><b>1) BeH</b> Die Berufsbezeichnung der im Meldezeitraum von der Beschäftigten oder dem Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für eine Beschäftigte oder 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 Beschäftigten oder 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 (siehe Abschnitt 2.3). 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	<p><b>1) BeH</b> In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf.</p> <p><b>2) LeH</b> Das Merkmal wurde früher für die Quelle LeH nicht gefüllt, da die Berufsangaben in der LeH eine schlechte Qualität aufwiesen. Mittlerweile wird der Beruf in der LeH aus der Arbeitsuchendenhistorik zugespielt, so dass die Qualität gestiegen ist.</p> <p><b>3) XASU, XMTH</b> Das Merkmal ist für den gesamten zur Verfügung stehenden Zeitraum sehr schlecht gefüllt.</p>

#### 6.5.4 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, XMTH
Datentyp	Numerisch
Detailbeschreibung	<p><b>1) BeH</b> Die Berufsbezeichnung der im Meldezeitraum von der Beschäftigten oder dem 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 oder 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.</p>

	<p>derungsniveau 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, XMTH</b></p> <p>Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1).</p>
Qualitätshinweis	<p><b>1) BeH</b></p> <p>In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf.</p> <p><b>2) LeH</b></p> <p>Das Merkmal wurde früher für die Quelle LeH nicht gefüllt, da die Berufsangaben in der LeH eine schlechte Qualität aufwiesen. Mittlerweile wird der Beruf in der LeH aus der Arbeitssuchendenhistorik zugespielt, so dass die Qualität gestiegen ist.</p> <p><b>3) XASU, XMTH</b></p> <p>Das Merkmal ist für den gesamten zur Verfügung stehenden Zeitraum sehr schlecht gefüllt.</p>
Besonderheiten	<p>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.</p>

### 6.5.5 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, XMTH
Datentyp	Numerisch
Detaillbeschreibung	<p><b>1) BeH</b></p> <p>Die Berufsbezeichnung der im Meldezeitraum von der oder dem 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 oder 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. November 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></p> <p>Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1).</p>

Qualitätshinweis	<p><b>1) BeH</b> In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf.</p> <p><b>2) LeH</b> Das Merkmal wurde früher für die Quelle LeH nicht gefüllt, da die Berufsangaben in der LeH eine schlechte Qualität aufwiesen. Mittlerweile wird der Beruf in der LeH aus der Arbeitssuchendenhistorik zugespielt, so dass die Qualität gestiegen ist.</p> <p><b>3) XASU, XMTH</b> Das Merkmal ist für den gesamten zur Verfügung stehenden Zeitraum sehr schlecht gefüllt.</p>
------------------	--

### 6.5.6 Teilzeit (teilzeit)

Variablenlabel	Teilzeit
Variablenname	teilzeit
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal Teilzeit (teilzeit) unterscheidet zwischen Vollzeit- und Teilzeitbeschäftigten. Maßgeblich ist dabei das Verhältnis zwischen der arbeitsvertraglich vereinbarten und der betriebsüblichen Arbeitszeit.
Qualitätshinweis	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. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf. Hier wird keine Imputation vorgenommen.

### 6.5.7 Stellung im Beruf und Arbeitszeit (stib)

Variablenlabel	Stellung im Beruf und Arbeitszeit
Variablenname	stib
Kategorie	Informationen zu Beschäftigung
Herkunft	BeH
Datentyp	Numerisch
Detailbeschreibung	<p>Die berufliche Stellung eines Beschäftigten im Meldezeitraum wird vom Arbeitgeber im Rahmen der „Angaben zur Tätigkeit“ übermittelt. Das Merkmal „Stellung im Beruf und Arbeitszeit“ 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. Die Unterscheidung zwischen vollzeitbeschäftigten Arbeitern und Angestellten richtet sich ausschließlich nach der Art der Rentenversicherungspflicht (BfA für Angestellte und LVA für Arbeiter). Meister und Poliere werden nur dann in einer eigenen Kategorie erfasst, wenn sie in der Arbeiterrentenversicherung pflichtversichert sind. Die Zuordnung eines Meisters</p>

	<p>oder Poliers zu den Arbeitern oder Angestellten kann nur über den jeweiligen Rententräger erfolgen. Neben den Auszubildenden, Praktikanten und Volontären zählen auch Anlernlinge, Schüler an Schulen des Gesundheitswesens und Teilnehmer an geförderten Maßnahmen zur beruflichen Fortbildung, Umschulung und betrieblichen Einarbeitung zur Kategorie Beschäftigte in beruflicher Ausbildung.</p> <p>Falls mehrere Schlüsselzahlen in Betracht kommen, ist der Arbeitgeber dazu angehalten, die Einstufung nach der überwiegend ausgeübten Tätigkeit vorzunehmen. Lässt sich diese nicht eindeutig feststellen, ist die Schlüsselzahl der höherwertigen Stellung im Beruf einzutragen (vgl. Bundesagentur für Arbeit 2005, S. VI).</p> <p>Mit der Einführung des neuen Tätigkeitsschlüssels (siehe Kapitel 2.3) fiel diese Unterscheidung jedoch weg. Die Variable „stib“ ist daher nur für Meldungen gefüllt, die noch vor Einführung des neuen Tätigkeitsschlüssels abgegeben wurden.</p>
Qualitätshinweis	In 1991 kommt es aufgrund der Wiedervereinigung zu einer deutlichen Häufung der fehlenden Angaben.

### 6.5.8 Erwerbsstatus (erwstat)

Variablenlabel	Erwerbsstatus
Variablenname	erwstat
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
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 01.01.1999 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 es sich dabei um sozialversicherungspflichtige Beschäftigungsverhältnisse ohne Besonderheiten, die unter der 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.</p> <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.</p> <p>Seit dem 01.04.1999 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.</p> <p><b>2) LeH</b></p> <p>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.</p> <p><b>3) LHG</b></p> <p>Für LHG-Sätze zeigt das Merkmal Erwerbsstatus an, ob die Person minderjährig er-</p>

	<p>werbsfähig, volljährig erwerbsfähig oder nicht erwerbsfähig ab Altersrentengrenze geführt wird.</p> <p>Da die Meldelogik in vielen Fällen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben durch das in Abschnitt 3.1.9 beschriebene Anonymisierungsverfahren verändert.</p> <p><b>4) ASU, XASU</b></p> <p>Für ASU/XASU-Sätze zeigt das Merkmal Erwerbsstatus den Arbeitsvermittlungsstatus an. Es wird zwischen „arbeitslos arbeitssuchend“, „nicht arbeitslos arbeitssuchend“, „ratsuchend“ und „nicht arbeitssuchend“ unterschieden. Als Ratsuchende gelten Bewerber, die von der BA nur eine Beratung wünschen. Dazu kommen Rehabilitanden und vor 2008 Personen ab dem 58. Lebensjahr, die einer Vermittlung nicht uneingeschränkt zur Verfügung stehen. Ab 01.08.2016 können auch Personen dazu zählen, deren Hilfebedürftigkeit nicht mehr besteht, die aber dennoch weiter vom Jobcenter betreut werden.</p> <p>„Nicht arbeitssuchend“ 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.</p> <p>Da die Meldelogik in vielen Fällen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben durch das im Abschnitt 3.1.9 beschriebene Anonymisierungsverfahren verändert.</p> <p><b>5) MTH, XMTH</b></p> <p>Für MTH- und XMTH-Sätze zeigt das Merkmal Erwerbsstatus die Maßnahmeart-Kategorie an. Dies ist die höchste Hierarchiestufe der Maßnahmeart-Systematiken der Bundesagentur für Arbeit.</p>
Qualitätshinweis	<p><b>1) LHG</b></p> <p>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.</p> <p>Es kann in Einzelfällen zu widersprüchlichen Informationen zur Erwerbsfähigkeit einer Person geben, falls diese sich zeitgleich in verschiedenen Bedarfsgemeinschaften (BGs) befindet. Ein möglicher Grund dafür ist die Feststellung der Erwerbsfähigkeit durch die die BG betreuende Jobcenter-Fachkraft bzw. die jeweils zuständigen medizinischen Dienste. Sie können zu unterschiedlichen Ergebnissen kommen oder die Ergebnisse zu unterschiedlichen Zeitpunkten dokumentieren. Ein weiterer Grund können unterschiedliche Geburtstagsangaben für die zeitgleichen BG-Kunden sein, die sich auf den Erwerbsfähigkeitsstatus auswirken können.</p> <p><b>2) ASU, XASU</b></p> <p>Die Kategorien „ratsuchend“ und „nicht arbeitssuchend“ sind erst seit der Einführung von Verbis (siehe Kapitel 2.6) vorhanden. Aufgrund von Nachmeldungen sind aber auch Eintragungen vor 2006 zu finden. Das Merkmal „nicht arbeitssuchend“ gilt vor 2008 als untererfasst. In der XASU ist die Qualität der Statusbestimmung bei etlichen Trägern zum Teil eingeschränkt, insbesondere in den ersten Jahren.</p>

### 6.5.9 Gleitzone (gleitz)

Variablenlabel	Gleitzone
Variablenname	gleitz
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH

Datentyp	Numerisch
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 01.04.2003.

#### 6.5.10 Leiharbeit (leih)

Variablenlabel	Leiharbeit
Variablenname	leih
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
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. Im Jahr 2012 beträgt die Missingquote nur noch rund 3 %.

#### 6.5.11 Befristung (befrist)

Variablenlabel	Befristung
Variablenname	befrist
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
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. Im Jahr 2012 beträgt die Missingquote nur noch rund 3 %.

#### 6.5.12 Grund Abgabe Beschäftigungsmeldung / Grund Ende Leistungsbezug (grund)

Variablenlabel	Abmeldegrund / Abgabegrund / Beendigungsgrund
Variablenname	grund
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, LHG, ASU, XASU
Datentyp	Numerisch
Detailbeschreibung	<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 Meldeverfahren-

	<p>ens möglichen 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 darauffolgenden Jahres-, Unterbrechungs- oder Abmeldung erneut übermittelt werden.</p> <p>Die Abgabegründe für Beschäftigungsmeldungen sind nach den Regeln des seit 01.01.1999 geltenden Meldeverfahrens (gemäß DEÜV) kodiert.</p> <p><b>2) LeH</b></p> <p>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 bezieht sich auf die Person, 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></p> <p>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.</p> <p>Ab dem 26.04.2003 erfolgte eine Reduktion der Merkmalsausprägungen. Für Auswertungen über längere Zeiträume können die alten Werte auf die aktuell gültigen anhand folgender Tabelle umgeschlüsselt werden (im 5000er Zahlenbereich):</p> <table border="1" data-bbox="534 1249 1460 1400"> <tr> <td>29→60</td> <td>33→60</td> <td>37→66</td> <td>42→65</td> <td>46→67</td> <td>50→75</td> <td>54→78</td> </tr> <tr> <td>30→60</td> <td>34→60</td> <td>38→66</td> <td>43→70</td> <td>47→67</td> <td>51→74</td> <td></td> </tr> <tr> <td>31→61</td> <td>35→60</td> <td>39→71</td> <td>44→74</td> <td>48→78</td> <td>52→76</td> <td></td> </tr> <tr> <td>32→60</td> <td>36→61</td> <td>40→69</td> <td>45→77</td> <td>49→69</td> <td>53→68</td> <td></td> </tr> </table> <p><b>5) XASU</b></p> <p>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>	29→60	33→60	37→66	42→65	46→67	50→75	54→78	30→60	34→60	38→66	43→70	47→67	51→74		31→61	35→60	39→71	44→74	48→78	52→76		32→60	36→61	40→69	45→77	49→69	53→68	
29→60	33→60	37→66	42→65	46→67	50→75	54→78																							
30→60	34→60	38→66	43→70	47→67	51→74																								
31→61	35→60	39→71	44→74	48→78	52→76																								
32→60	36→61	40→69	45→77	49→69	53→68																								
Qualitätshinweis	<p><b>1) BeH</b></p> <p>Ab 2013 nimmt die Anzahl der Meldungen mit Abmeldegrund 54 (Meldung eines einmalig gezahlten Entgelts) sprunghaft zu. Solange ein Beschäftigungsverhältnis besteht, können Sonderzahlungen, die bis März des Folgejahres ausgezahlt werden, mit der normalen Meldung (v.a. Jahresmeldung) für das Vorjahr angegeben werden. Eine gesonderte Meldung mit Abmeldegrund 54 ist dann nicht erforderlich. Bis 2012 konnten die Jahresmeldungen bis Mitte April abgegeben werden, seit 2013 müssen sie nun bis spätestens Mitte Februar erfolgen. Für Sonderzahlungen, die im Februar und März ausgezahlt werden, muss nun eine extra Meldung erfolgen.</p> <p><b>2) LHG, XASU</b></p> <p>Der Anteil valider Werte (Füllgrad) des Abgabegrundes ist in der Quelle LHG über alle Jahre sehr niedrig (&lt; 20%).</p>																												

### 6.5.13 Erwerbsstatus vor Arbeitsuche (estatvor)

Variablenlabel	Erwerbsstatus vor Arbeitsuche
Variablenname	estatvor
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, XASU
Datentyp	Numerisch
Detailbeschreibung	<p>Die Variable zeigt für ASU- und XASU-Sätze den Erwerbsstatus vor der Arbeitsuche an. Ab Dezember 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.</p> <p>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.</p>

### 6.5.14 Status nach Arbeitsuche (estatnach)

Variablenlabel	Status nach Arbeitsuche
Variablenname	estatnach
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU
Datentyp	Numerisch
Detailbeschreibung	<p>Das Merkmal enthält den Status der Person nach Abgang aus der Arbeitslosigkeit. Längere Krankheitsepisoden können über dieses Merkmal identifiziert werden. 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>

### 6.5.15 Profillage (profil)

Variablenlabel	Profillage
Variablenname	profil
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, MTH
Datentyp	numerisch
Detailbeschreibung	<p>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 Profiling wird zusammenfassend eine Profillage festgestellt. Hierzu wird zunächst die allgemeine Integrationsprognose des Kunden bestimmt. Zur Auswahl stehen:</p> <ul style="list-style-type: none"> <li>• integrationsnah (Integration in den 1. Arbeitsmarkt innerhalb von 12 Monaten realistisch)</li> </ul>

	<ul style="list-style-type: none"> <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.

#### 6.5.16 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
Detailbeschreibung	Dieses Merkmal beschreibt die Art der Beendigung des letzten Arbeits- oder Ausbildungsverhältnisses vor einer Arbeitsuche. Es kann daher zur Identifikation von Job-to-Job-Vermittlungen verwendet werden.

#### 6.5.17 Arbeitszeit des Stellengesuchs (arbeit)

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

#### 6.5.18 Restanspruch / geplante Dauer (restanspruch)

Variablenlabel	Restanspruch / geplante Dauer
Variablenname	restanspruch
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LeH, MTH
Datentyp	Numerisch
Detailbeschreibung	<p>Je nach Datenquelle hat das Merkmal eine unterschiedliche Bedeutung.</p> <p><b>1) LeH</b></p> <p>Das Merkmal enthält den Anspruch, der nach Ende des aktuellen Leistungsbezugs verbleibt. 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. Aller-</p>

	<p>dings 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.</p> <p>Liegt das Enddatum des Leistungsempfangs vor dem 01.01.1998, wird der Restanspruch in Werktagen angegeben. Danach erfolgt die Angabe in Kalendertagen.</p> <p>Die Angabe bezieht sich nicht auf das Enddatum der Episode, sondern auf das Ende des Originalzeitraums.</p> <p><b>2) MTH</b></p> <p>Das Merkmal enthält die geplante Dauer der Maßnahme.</p>
Qualitätshinweis	<p><b>1) LeH</b></p> <p>Im Verlauf einer Leistungsbiografie ist grundsätzlich mit einer systematischen Entwicklung des Restanspruchs zu rechnen. Das bedeutet, dass der Restanspruch am Ende einer Leistungsperiode minus der Dauer der Folgeperiode den Restanspruch nach der Folgeperiode ergeben sollte, wenn in der Zwischenzeit keine Ansprüche entstanden oder verfallen sind. Es kommt jedoch teilweise zu unerwarteten Erhöhungen oder Reduzierungen des Restanspruchs.</p> <p>Hierbei handelt es sich vermutlich überwiegend um Korrekturen bei der Bearbeitung von Leistungsfällen. Solche Korrekturen treten zum Beispiel auf, wenn ein Anspruch zu Beginn des Leistungsfalls aufgrund unvollständiger Informationen nicht korrekt festgelegt oder ein noch nicht verfallener Restanspruch nicht berücksichtigt wurde. Diese Korrekturen erfolgen dann aber in der Regel nur in der aktuellen Buchung, eine Bereinigung des kompletten Falls findet nicht statt. Darüber hinaus kann die Angabe zur Anspruchsdauer in Einzelfällen die aus den gesetzlichen Vorgaben abzuleitende individuelle Obergrenze übersteigen.</p>

#### 6.5.19 Trägerart (traeger)

Variablenlabel	Trägerart
Variablenname	traeger
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	<p>Das Merkmal enthält die Trägerschaftsart des Trägers der Grundsicherung (LHG), des Trägers der Maßnahme (MTH/XMTH) 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 Kapitel 2.5.</p> <p>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</p>
	<p>mit den Grenzen der Kreise übereinstimmen müssen. Gleiches gilt für die Arbeitsagenturen.</p> <p>Bei ASU- und XASU-Sätzen ist die Art des Trägers enthalten, der das Bewerberangebot führt.</p>

#### 6.5.20 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, XMTH

Datentyp	Numerisch
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</li> </ul>

### 6.5.21 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, XMTH
Datentyp	Numerisch
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</li> </ul> <p>Bei der Berechnung der Dauer werden diese Lücken jedoch nicht addiert. Daher kann sich die Dauer in diesem Merkmal von dem Ergebnis der Dauerberechnung mit dem "Beginndatum der Arbeitslosigkeit" unterscheiden.</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

### 6.6.1 Wohnort – Kreis (wo\_kreis)

Variablenlabel	Wohnort - Kreis
Variablenname	wo_kreis
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
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 (NUTS-1) und in der ersten bis dritten Stelle den Regierungsbezirk (NUTS-2). Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null. In der Systematik NUTS (Nomenclature des unités territoriales statistiques) der Europäischen Union entsprechen die Kreise der Hierarchieebene NUTS 3.</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 gesamten Zeitraum der Originalbeobachtung. Bei ASU, LeH, MTH und XMTH enthält das Merkmal den Wohnort zum Beginn des Originalzeitraums. Mit der Dauer des Spells steigt also das Risiko, dass der Wohnort veraltet ist und die Angaben zu</p>

	<p>späteren Stichtagen falsch sind.</p> <p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben für alle Quellen auf den Gebietsstand vom 31.12.2017 umgeschlüsselt, d. h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2017 hatten. Da sich die Grenzen der Kreise im Zeitverlauf geändert haben, würden ohne Gebietsstandaktualisierung Fälle auftreten, bei denen die Kreiskennziffer des Wohnortes wechselt, ohne dass der / die betreffende Versicherte umgezogen ist.</p>
Qualitätshinweis	<p>Hinsichtlich der Wohnortangaben bestehen für einzelne Beschäftigte Erhebungsungenauigkeiten. Die Meldevorschrift stellt nicht klar, welcher Wohnsitz – Haupt- oder Nebenwohnsitz mit überwiegender Aufenthaltsort – vom Arbeitgeber zu melden ist. Dies kann in der Beschäftigungsstatistik zum Nachweis von "Fernpendlern" zwischen gemeldetem Hauptwohnsitz und Arbeitsort führen, obwohl die beschäftigte Person ihrer Beschäftigung am Nebenwohnsitz nachgeht, also faktisch nicht pendelt.</p> <p>Im Jahr 2015 kommt es in den Daten zu einer Reduktion der Wohnortwechsel von ca. 10-15% gegenüber dem üblichen Niveau der umliegenden Jahre, wobei regionale Unterschiede auftreten. Der genaue Grund für diese Abweichung konnte bisher nicht ermittelt werden.</p>
Besonderheiten	<p>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.</p>

### 6.6.2 Wohnort – Bundesland (wo\_bula)

Variablenlabel	Wohnort - Bundesland
Variablenname	wo_bula
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
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>

### 6.6.3 Wohnort – Arbeitsagentur (wo\_aa)

Variablenlabel	Wohnort - Arbeitsagentur
Variablenname	wo_aa
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Hierarchie	Regionaldirektion   Arbeitsagentur
Detailbeschreibung	<p>Das Merkmal enthält ab 1999 den Agenturbezirk der am Wohnort des Beschäftigten/Kunden zuständigen Arbeitsagentur. Sie wird aus der Wohnadresse ermittelt. Für die Quellen LHG und XASU gilt der Wohnort für den Zeitraum der Originalbeobachtung. Bei LeH, ASU, MTH und XMTH enthält das Merkmal den Wohnort zum</p>

	<p>Beginn der Meldung. Mit der Dauer des Spells steigt also das Risiko, dass der Wohnort veraltet ist und die Angaben zu späteren Stichtagen falsch sind. Es ist für die BeH und LeH erst ab 1999 vorhanden.</p> <p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Angaben für alle Quellen auf den Gebietsstand vom 31.12.2017 umgeschlüsselt, d. h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Agenturbezirk nach den Grenzen, welche die Bezirke zum 31.12.2017 hatten. Da sich die Grenzen im Zeitverlauf geändert haben, würden ohne diese Gebietsstandaktualisierung Fälle auftreten, bei denen der Bezirk einer Person wechselt, ohne dass diese Person umgezogen ist. Einen Problemfall stellt hinsichtlich der Gebietsstandaktualisierung allerdings Berlin dar: Die Berliner Arbeitsagenturbezirke wurden im Laufe der Jahre immer wieder neu zugeschnitten, was auch durch die Gebietsstandumschlüsselung nicht vollständig bereinigt werden konnte. Berlin sollte möglichst als Einheit und nicht getrennt nach Ost und West analysiert werden, da seit der Gebietsreform der Arbeitsämter in Berlin zum 01.07.1997 deren eindeutige Zuordnung zu Ost bzw. West nicht mehr möglich ist. Von der Statistik der BA wird Berlin einheitlich dem Osten zugeordnet.</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 Gebietsstandaktualisierung nicht möglich. Dies ist bei Berechnungen im Zeitverlauf (z. B. regionale Arbeitslosenzahlen) zu beachten.</p>
Besonderheiten	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.

#### 6.6.4 Wohnort – Regionaldirektion (wo\_rd)

Variablenlabel	Wohnort - Regionaldirektion
Variablenname	wo_rd
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
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

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

Variablenlabel	WZ73 3-Steller
Variablenname	w73_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
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.</p> <p>Jedem Betrieb ist nur eine Kennziffer zugeordnet. Die Zuordnung zu der entsprechenden Wirtschaftsklasse erfolgt unter Beachtung der institutionellen Ausrichtung des Betriebes.</p>
--------------------	--

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

Variablenlabel	WZ93 5-Steller
Variablenname	w93_5
Kategorie	Betriebsmerkmale
Herkunft	BHP
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 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 (w93_3) bereitgestellt.

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

Variablenlabel	WZ93 3-Steller
Variablenname	w93_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
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 widerspiegelt werden.</p>

	<p>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 widerspiegelt werden.</p>
--	--

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

Variablenlabel	WZ03 5-Steller
Variablenname	w03_5
Kategorie	Betriebsmerkmale
Herkunft	BHP
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 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>
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.

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

Variablenlabel	WZ03 3-Steller
Variablenname	w03_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
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 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>

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

Variablenlabel	WZ08 5-Steller
Variablenname	w08_5
Kategorie	Betriebsmerkmale
Herkunft	BHP
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.

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

Variablenlabel	WZ08 3-Steller
Variablenname	w08_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
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>

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

Variablenlabel	w73_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w73_3_gen
Kategorie	Betriebsmerkmale
Herkunft	BHP
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>
--------------------	--

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

Variablenlabel	Art der Vervollständigung w73_3
Variablenname	group_w73_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
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>

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

Variablenlabel	W93_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w93_3_gen
Kategorie	Betriebsmerkmale
Herkunft	BHP
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- oder 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>

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

Variablenlabel	Art der Vervollständigung w93_3
Variablenname	group_w93_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch

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>
--------------------	--

#### 6.7.12 w08\_3 vervollständigt durch Extrapolation/Imputation (w08\_3\_gen)

Variablenlabel	w08_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w08_3_gen
Kategorie	Betriebsmerkmale
Herkunft	BHP
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. Von 2008 bis 2017 enthält das Merkmal die Originalwerte aus w08_3. Vor 2008 wird die Information rückgeschrieben oder mit Hilfe von Umschlüsselungstabellen ersetzt, sodass das Merkmal eine zeitkonsistente Wirtschaftszweigangabe auf Basis der Wirtschaftszweigklassifikation WZ08 enthält. Eine ausführliche Beschreibung dazu findet sich bei Eberle et al., 2011.</p> <p>Weitere Informationen zur Klassifikation WZ08 finden sich bei der Variablenbeschreibung w08_3.</p>

#### 6.7.13 Art der Vervollständigung w08\_3 (group\_w08\_3)

Variablenlabel	Art der Vervollständigung w08_3
Variablenname	group_w08_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	<p>Dieses Merkmal weist die Art der Vervollständigung für das Merkmal w08_3_gen aus. Es ist angegeben, ob der jeweilige Wert in w08_3_gen dem Originalwert von w08_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>

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

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

	<p>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.</p> <p>Ein Betrieb muss nicht zwingend in dem Jahr des ersten Auftretens im BHP enthalten sein, da für das BHP nur der Stichtag 30.06. relevant ist. Hat der Betrieb in seinem Gründungsjahr an diesem Stichtag keine Beschäftigte, erscheint er auch nicht im BHP.</p>
--	---

#### 6.7.15 Erstes Auftreten Betriebsnummer (grd\_dat)

Variablenlabel	Erstes Auftreten Betriebsnummer
Variablenname	grd_dat
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Datum
Detailbeschreibung	<p>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.</p> <p>Ein Betrieb muss nicht zwingend in dem Jahr des ersten Auftretens im BHP enthalten sein, da für das BHP nur der Stichtag 30.06. relevant ist. Hat der Betrieb in seinem Gründungsjahr an diesem Stichtag keine Beschäftigte, erscheint er auch nicht im BHP.</p>
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.

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

Variablenlabel	Jahr des letzten Auftretens der Betriebsnummer
Variablenname	lzt_jahr
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	<p>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 Betriebsstilllegung 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 neuen Nummer“ oder andere verwaltungstechnische Änderungen in Betracht (vgl. Bender et al., 1996 oder Bundesagentur für Arbeit, 2007, S. 9-11).</p>

	Ein Betrieb muss nicht zwingend in dem Jahr des letzten Auftretens im BHP enthalten sein, da für das BHP nur der Stichtag 30.06. relevant ist. Hat der Betrieb in seinem Schließungsjahr an diesem Stichtag keine Beschäftigte, erscheint er auch nicht im BHP.
--	---

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

Variablenlabel	Letztes Auftreten Betriebsnummer
Variablenname	lzt_dat
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Datum
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 neuen Nummer“ oder andere verwaltungstechnische Änderungen in Betracht (vgl. Bender et al., 1996 oder Bundesagentur für Arbeit, 2007, S. 9-11). Ein Betrieb muss nicht zwingend in dem Jahr des letzten Auftretens im BHP enthalten sein, da für das BHP nur der Stichtag 30.06. relevant ist. Hat der Betrieb in seinem Schließungsjahr an diesem Stichtag keine Beschäftigte, erscheint er auch nicht im BHP.
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.

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

Variablenlabel	Anzahl Beschäftigte gesamt
Variablenname	az_ges
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
Detailbeschreibung	Die Variable enthält die Anzahl aller zum 30.06. eines Jahres zur Sozialversicherung gemeldeten Mitarbeiter eines Betriebes. Seit Einführung der neuen Meldevorschriften im Jahr 1999 sind auch die geringfügig Beschäftigten erfasst. Ruhende Beschäftigungsverhältnisse (Tagesentgelt von Null) sind nicht enthalten.

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

Variablenlabel	Anzahl Vollzeit (Normalbeschäftigte + sonstige)
Variablenname	az_vz
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
Detailbeschreibung	Die Variable enthält die Anzahl der Personen im Betrieb, die die am 30.06. eines Jahres mit den Personengruppenschlüsseln 101, 140, 143, 105, 106, 112, 113, 114, 118, 119, 120, 149, 201, 203, 205, 999 und YYY und als Vollzeitbeschäftigte gemeldet sind. Azubis, geringfügig Beschäftigte und Personen in Altersteilzeit werden nicht berücksichtigt.

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

Variablenlabel	Anzahl geringfügig Beschäftigte
Variablenname	az_gf
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
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.

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

Variablenlabel	Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte
Variablenname	te_imp_mw
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
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 3.1.3.3 in Schmucker et al. 2018). Diese Daten wurden dann auf Betriebsebene aggregiert. Die Werte sind auf zwei Dezimalstellen gerundet. Aufgrund des "storage types" in Stata werden jedoch weitere Dezimalstellen angezeigt, die nicht korrekt sind.</p>

### 6.7.22 Arbeitsort Kreis (ao\_kreis)

Variablenlabel	Arbeitsort Kreis
Variablenname	ao_kreis
Kategorie	Ortsangaben
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Bundesland   Kreis
Detailbeschreibung	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 Bundeslandes (NUTS1), in der ersten bis dritten Stelle den Regierungsbezirk (NUTS2). Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null. In der Systematik NUTS (Nomenclature des

	<p>unités territoriales statistiques) der Europäischen Union entsprechen die Kreise der Hierarchieebene NUTS 3.</p> <p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben auf den Gebietsstand vom 31.12.2017 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Betriebsortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2017 hatten. Da sich die Grenzen der Kreise im Zeitverlauf geändert haben, würden ohne Gebietsstandaktualisierung 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.

### 6.7.23 Arbeitsort Bundesland (ao\_bula)

Variablenlabel	Arbeitsort Bundesland
Variablenname	ao_bula
Kategorie	Ortsangaben
Herkunft	BHP
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 (NUTS1).

## 6.8 Verknüpfungsmerkmale

### 6.8.1 Vorliegen administrativer Personendaten (match\_admin)

Variablenlabel	Vorliegen administrativer Personendaten
Variablenname	match_admin
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	Generiert
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal ist nur in der Verknüpfungsdatei enthalten. Es zeigt für jede befragte Person in den NEPS-SC5-Erhebungsdaten an, ob zu dieser Person auch Beobachtungen aus den administrativen Personendaten vorliegen. Diese Variable nimmt in wenigen Fällen den Wert 0 an, obwohl die Variable match_typ einen erfolgreichen Match anzeigt. In diesen Fällen konnte zwar ein Match durch Abgleich der Identifikatoren beider Datenquellen erreicht werden, die IEB enthält jedoch zu dieser Person keine Episoden.

### 6.8.2 Verknüpfungsmethode (match\_typ)

Variablenlabel	Verknüpfungsart
Variablenname	match_typ
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	Record Linkage
Datentyp	Numerisch

Detailbeschreibung	Dieses Merkmal ist nur in der Verknüpfungsdatei enthalten. Es beschreibt, auf welcher Stufe des Verknüpfungsprozesses es gelungen ist, eine Person in den administrativen Daten des IAB zu identifizieren.
--------------------	--

### 6.8.3 Qualität der Verknüpfung (quality)

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

## 7 Literatur

- Antoni, M., Schmucker, A., Seth, S., & Berge, P. v. (2019). Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975-2017. FDZ-Datenreport 02/2019 (de), Nürnberg.
- Antoni, M., & Schnell, R. (2019). The past, present and future of the German Record Linkage Center (GRLC). In: Jahrbücher für Nationalökonomie und Statistik. Journal of Economics and Statistics, De Gruyter, Jg. 239 (2), S. 1-13.
- Bender, S., Hilzendege, J., Rohwer, G., & Rudolph, H. (1996). Die IAB-Beschäftigtenstichprobe 1975-1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197, Nürnberg.
- Bertat, T., Dundler, A., Grimm, C., Kiewitt, J., Schomaker, C., Schridde, H., & Zemann, C. (2013). Neue Erhebungsinhalte „Arbeitszeit“, „ausgeübte Tätigkeit“ sowie „Schul- und Berufsabschluss“ in der Beschäftigungsstatistik. Methodenbericht der Statistik der Bundesagentur für Arbeit.
- Bundesagentur für Arbeit (2005). Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung (Ausgabe Januar 2005), Nürnberg.
- Bundesagentur für Arbeit (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. Methodenbericht der Statistik der Bundesagentur für Arbeit, Nürnberg.
- Bundesagentur für Arbeit (2011). Klassifikation der Berufe 2010 – Band 1: systematischer und alphabetischer Teil mit Erläuterungen. Klassifikation der Berufe 2010, Nürnberg.
- Bundesanstalt für Arbeit (1988). Klassifikation der Berufe – Systematisches und Alphabetisches Verzeichnis der Berufsbenennung, Nürnberg.
- Christen, P. (2012). Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection, Berlin, Springer.
- Cramer, U. (1985). Probleme der Genauigkeit der Beschäftigtenstatistik. Allgemeines Statistisches Archiv 69, S. 56-68.
- Eberle, J., Jacobebbinghaus, P., Ludsteck, J., & Witter, J. (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, B., Osikominu, A., & Völter, R. (2006). Imputation rules to improve the education variable in the IAB employment subsample. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 126 (3), S. 405-436.
- Ludsteck, J., & Thomsen, U. (2016). Imputation of the working time information for the employment register data. FDZ-Methodenreport 01/2016 (en), Nürnberg.
- Meinken, H., & Koch, I. (2004). BA-Beschäftigtenpanel 1998-2002. Codebuch, Nürnberg.
- Paulus, W., & Matthes, B. (2013). Klassifikation der Berufe: Struktur, Codierung und Umsteigeschlüssel. FDZ-Methodenreport 08/2013 (de), Nürnberg.
- Schild, C.-J., & Antoni, M. (2014). Linking survey data with administrative social security data – The project „Interactions between Capabilities in Work and Private Life“. German RLC Working Paper Nr. 2 WP-GRLC-2014-02.

- Schmucker, A., Eberle, J., Eberle, J., Ganzer, A., Steigmaier, J., & Umkehrer, M. (2018). Betriebs-Historik-Panel 1975-2016. FDZ-Datenreport 01/2018 (de), Nürnberg.
- Schnell, R., Bachteler, T., & Bender, S. (2004). A toolbox for record linkage. *Austrian Journal of Statistics*, Jg. 33(1-2), S. 125-133.
- Statistisches Bundesamt (2002). Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ93), Wiesbaden. URL: <https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-1993.html> (Zugriff: 28.07.2020).
- Statistisches Bundesamt (2003). Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003), Wiesbaden. URL: <https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2003.html?nn=205976> (Zugriff: 28.07.2020).
- Statistisches Bundesamt (2008). Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), Wiesbaden. URL: <https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html> (Zugriff: 28.07.2020).
- Statistisches Bundesamt (2020). Staats- und Gebietssystematik, Wiesbaden. URL: [https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluesel\\_pdf.pdf?blob=publicationFile&v=5%20](https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluesel_pdf.pdf?blob=publicationFile&v=5%20) (Zugriff: 28.07.2020).
- Thomsen, U., Ludsteck, J., & Schmucker, A. (2018). Skilled or unskilled – Improving the information on qualification for employee data in the IAB Employee Biography. FDZ-Methodenreport 09/2018 (en), Nürnberg.
- Wermter, W., & Cramer, U. (1988). Wie hoch war der Beschäftigtenanstieg seit 1983? – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, Jg. 21(4), 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://fdz.iab.de>.

## 8.2 Abkürzungsverzeichnis

<b>ALG II</b>	Arbeitslosengeld II
<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>BG</b>	Bedarfsgemeinschaft
<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>DZHW</b>	Deutsches Zentrum für Hochschul- und Wissenschaftsforschung
<b>EDV</b>	Elektronische Datenverarbeitung
<b>FDZ</b>	Forschungsdatenzentrum
<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>IAB-FDZ</b>	Forschungsdatenzentrum der Bundesagentur für Arbeit am IAB
<b>IEB</b>	Integrierte Erwerbsbiografien

<b>infas</b>	Institut für angewandte Sozialwissenschaften GmbH
<b>ISIC</b>	International Standard Industrial Classification of All Economic Activities
<b>LeH</b>	Leistungsempfängerhistorik
<b>LHG</b>	Leistungshistorik Grundsicherung
<b>LifBi</b>	Leibniz Institut für Bildungsverläufe
<b>LifBi-FDZ</b>	Forschungsdatenzentrum des Leibniz Institut für Bildungsverläufe
<b>LVA</b>	Landesversicherungsanstalt
<b>MTH</b>	Maßnahmeteilnahmehistorik
<b>NACE</b>	Nomenclature générale des activités économiques dans les communautés européennes
<b>NEPS</b>	National Educational Panel Study
<b>NEPS-SC5</b>	NEPS Startkohorte 5
<b>NUTS</b>	Nomenclature des unités territoriales statistique
<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>XMTH</b>	Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II
<b>zkT</b>	Zugelassene kommunale Träger

# Impressum

## FDZ-Datenreport 10|2020

### Veröffentlichungsdatum

25. September 2020

### Herausgeber

Forschungsdatenzentrum  
der Bundesagentur für Arbeit  
im Institut für Arbeitsmarkt- und Berufsforschung  
Regensburger Straße 104  
90478 Nürnberg

### Rechte

Nachdruck - auch auszugsweise - nur mit Genehmigung des FDZ gestattet

### Bezugsmöglichkeit dieses Dokuments

[http://doku.iab.de/fdz/reporte/2020/DR\\_10-20.pdf](http://doku.iab.de/fdz/reporte/2020/DR_10-20.pdf)

### Dokumentation Version

NEPS-SC5-ADIAB 7518\_DE\_v1\_dok1, DOI: 10.5164/IAB.FDZD.2010.de.v1

### Datensatz Version

NEPS-SC5-ADIAB 7518 v1, DOI: 10.5164/IAB.NEPS-SC5-ADIAB7518.de.en.v1

### Bezugsmöglichkeit aller Veröffentlichungen der Reihe „FDZ-Datenreport“

[https://fdz.iab.de/de/FDZ\\_Publications/FDZ\\_Publication\\_Series/FDZ-Datenreporte.aspx](https://fdz.iab.de/de/FDZ_Publications/FDZ_Publication_Series/FDZ-Datenreporte.aspx)

### Website

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

---

### Rückfragen zum Inhalt

Clara Wolf

Telefon: 0911 179-6807

E-Mail: [Clara.Wolf@iab.de](mailto:Clara.Wolf@iab.de)