



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

FDZ-DATENREPORT

Dokumentation zu Arbeitsmarktdaten

02|2026 DE Befragungsdaten der SOEP-Core-, IAB-SOEP Migrationsstichprobe, IAB-BAMF-SOEP Befragung von Geflüchteten und SOEP-Innovationssample verknüpft mit administrativen Daten des IAB (SOEP-CMI-ADIAB) 1975-2023

Manfred Antoni, Mattis Beckmannshagen, Markus M. Grabka, Sekou Keita & Parvati Trübswetter



Bundesagentur für Arbeit

Befragungsdaten der SOEP-Core-, IAB-SOEP Migrationsstichprobe, IAB-BAMF-SOEP Befragung von Geflüchteten und SOEP- Innovationssample verknüpft mit administrativen Daten des IAB (SOEP-CMI-ADIAB) 1975-2023

Manfred Antoni (IAB), Mattis Beckmannshagen (DIW Berlin), Markus M. Grabka (DIW Berlin),
Sekou Keita (IAB) & Parvati Trübswetter (IAB)

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

1	Einleitung und Kurzbeschreibung	7
1.1	Einleitung	7
1.2	Datennutzung	10
1.2.1	Datenzugang	10
1.2.2	Datenhaltung	10
1.2.3	Sensible Merkmale und Zusatzvariablen	11
1.3	Änderungen zur Vorgängerversion	13
1.3.1	Änderungen in den Befragungsdaten	13
1.3.2	Merkmalskanon	14
1.3.3	Datendokumentation	15
1.4	Steckbrief	15
1.5	Mengengerüst	19
2	Datenquellen	20
2.1	Das Sozio-oekonomische Panel (SOEP)	20
2.2	Administrative Daten aus den Integrierten Erwerbsbiografien (IEB)	21
2.3	Beschäftigtenhistorik (BeH)	22
2.4	Leistungsempfängerhistorik (LeH)	23
2.5	Leistungshistorik Grundsicherung (LHG)	24
2.6	Arbeitsuchendenhistoriken (ASU und XASU)	24
2.7	Maßnahmeteilnahmehistoriken (MTH und XMTH)	25
3	Datenaufbereitung und Stichprobenziehung	25
3.1	Bereinigungen und Aufbereitungen	25
3.1.1	SOEP-Befragungsdaten	25
3.1.2	Administrative Daten aus den Integrierten Erwerbsbiographien (IEB)	25
3.1.3	Beschäftigtenhistorik (BeH)	26
3.1.4	Leistungsempfängerhistorik (LeH)	26
3.1.5	Leistungshistorik Grundsicherung (LHG)	26
3.1.6	Arbeitsuchendenhistoriken (ASU und XASU)	27
3.1.7	Maßnahmeteilnahmehistoriken (MTH)	28
3.1.8	Maßnahmeteilnahmehistoriken aus XSozial (XMTH)	28
3.1.9	SGB-II-Anonymisierung	29
3.2	Episodensplitting	29
3.3	Fehlende Werte	30
4	Datenverknüpfung	30
4.1	Informierte Zustimmung der Befragten zur Datenverknüpfung	30

4.2	Herkunft der Verknüpfungsmerkmale	32
4.3	Bereinigung und Standardisierung personenbezogener Merkmale	32
4.4	Exakter Abgleich der Verknüpfungsmerkmale	33
4.5	Fehlertoleranter Abgleich der Verknüpfungsmerkmale	33
4.6	Ergebnis der Datenverknüpfung	35
5	Datenqualität und Probleme	35
5.1	Sozio-oekonomisches Panel (SOEP)	35
5.2	Gesamte IEB	35
5.2.1	Lücken in den Erwerbsbiografien	35
5.2.2	Einführung des SGB II und spätere Trägerwechsel	37
5.3	Beschäftigtenhistorik (BeH).....	38
5.4	Leistungsempfängerhistorik (LeH).....	40
5.5	Leistungshistorik Grundsicherung (LHG)	41
5.6	Arbeitsuchendenhistoriken (ASU/XASU).....	42
5.6.1	ASU.....	42
5.6.2	XASU.....	43
5.7	Maßnahmeteilnahmehistoriken (MTH/XMTH).....	43
5.7.1	MTH.....	43
5.7.2	XMTH.....	44
6	Variablenbeschreibung.....	45
6.1	Identifikatoren	45
6.1.1	Personen-ID (pid)	45
6.1.2	Betriebs-ID (betnr_soepadiab)	45
6.2	Generierte technische Merkmale	46
6.2.1	Satzzähler pro Konto (spell)	46
6.2.2	Quelle des Satzes (quelle).....	47
6.2.3	Jahr (jahr)	47
6.3	Gültigkeitszeitraum	47
6.3.1	Beginndatum Originalbeobachtung (begorig).....	47
6.3.2	Enddatum Originalbeobachtung (endorig).....	48
6.3.3	Beginndatum Episode (begepi)	48
6.3.4	Enddatum Episode (endepi)	49
6.4	Informationen zur Person.....	49
6.4.1	Geschlecht (frau)	49
6.4.2	Geburtsjahr (gebjahr).....	49
6.4.3	Geburtsmonat (gebmon)	49
6.4.4	Staatsangehörigkeit (nation).....	50
6.4.5	Staatsangehörigkeit vergrößert (nation_gr).....	50
6.4.6	Familienstand (famst)	50

6.4.7	Kinderzahl (kind)	51
6.4.8	Ausbildung (ausbildung)	52
6.4.9	Ausbildung (imputiert) (ausbildung_imp)	54
6.4.10	Schulausbildung (schule)	55
6.5	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche	56
6.5.1	Tagesentgelt / täglicher Leistungssatz (tentgelt)	56
6.5.2	Tagesentgelt (inkl. Einmalzahlungen) (tentgelt_bonus)	57
6.5.3	Tagesentgelt (imputiert) (tentgelt_imp)	58
6.5.4	Beruf – ausgeübte / letzte Tätigkeit (KldB 1988) (beruf)	58
6.5.5	Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010), 3-Steller (beruf2010_3)	59
6.5.6	Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010), 4-Steller (beruf2010_4)	60
6.5.7	Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau)	62
6.5.8	Teilzeit (teilzeit)	63
6.5.9	Stellung im Beruf und Arbeitszeit (stib)	63
6.5.10	Erwerbsstatus (erwstat)	64
6.5.11	Maßnahmeart – Gruppe (mass)	66
6.5.12	Gleitzone (gleitz)	66
6.5.13	Leiharbeit (leih)	67
6.5.14	Befristung (befrist)	67
6.5.15	Abmeldegrund / Abgabegrund / Beendigungsgrund (grund)	68
6.5.16	Erwerbsstatus vor Arbeitsuche (estatvor)	70
6.5.17	Status nach Arbeitsuche / SGB-II-Ausschlussgrund / Verfügbarkeit (estatnach)	70
6.5.18	Integrationsprognose (ipo)	70
6.5.19	Art der Kündigung der letzten Tätigkeit (art_kuend)	71
6.5.20	Arbeitszeit des Stellengesuchs (arbzeit)	71
6.5.21	Restanspruch/geplante Dauer (restanspruch)	72
6.5.22	Trägerart (traeger)	73
6.5.23	Beginndatum der Arbeitslosigkeit (alo_beg)	73
6.5.24	Dauer der Arbeitslosigkeit (alo_dau)	73
6.6	Ortsangaben	74
6.6.1	Wohnort – Kreis (wo_kreis)	74
6.6.2	Wohnort – Bundesland (wo_bula)	75
6.6.3	Wohnort – Arbeitsagentur (wo_aa)	75
6.6.4	Wohnort – Regionaldirektion (wo_rd)	76
6.7	Betriebsmerkmale	76
6.7.1	WZ73 3-Steller (w73_3)	76
6.7.2	WZ93 5-Steller (w93_5)	77
6.7.3	WZ93 3-Steller (w93_3)	77
6.7.4	WZ03 5-Steller (w03_5)	78
6.7.5	WZ03 3-Steller (w03_3)	78
6.7.6	WZ08 5-Steller (w08_5)	79

6.7.7	WZ08 3-Steller (w08_3)	79
6.7.8	w73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen)	80
6.7.9	Art der Vervollständigung w73_3 (group_w73_3)	80
6.7.10	w93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen)	80
6.7.11	Art der Vervollständigung w93_3 (group_w93_3)	81
6.7.12	w08_3 vervollständigt durch Extrapolation/Imputation (w08_3_gen)	81
6.7.13	Art der Vervollständigung w08_3 (group_w08_3)	82
6.7.14	Jahr erstes Auftreten Betriebsnummer (grd_jahr)	82
6.7.15	Jahr letztes Auftreten Betriebsnummer (lzt_jahr)	82
6.7.16	Anzahl Beschäftigte gesamt (az_ges)	83
6.7.17	Anzahl Vollzeit (Normalbeschäftigte + sonstige) (az_vz)	83
6.7.18	Anzahl geringfügig Beschäftigte (az_gf)	83
6.7.19	Mittelwert imputiertes Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)	84
6.7.20	Arbeitsort - Kreis (ao_kreis)	84
6.7.21	Arbeitsort - Bundesland (ao_bula)	85
6.8	Verknüpfungsmerkmale	85
6.8.1	Vorliegen administrativer Personendaten (match_admin)	85
6.8.2	Verknüpfungsart (match_typ)	85
6.8.3	Teilpopulation (group)	86
7	Literatur	86
8	Anhang	90
8.1	Labels, Auszählungen und fehlende Werte	90
8.2	Abkürzungsverzeichnis	90

Zusammenfassung

Dieser Datenreport beschreibt die verknüpften Befragungsdaten der SOEP-Core-, IAB-SOEP Migrationsstichprobe, IAB-BAMF-SOEP Befragung von Geflüchteten und SOEP-Innovationsstichprobe mit administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB).

Abstract

This data report describes the linked survey data of SOEP Core, IAB-SOEP Migration Sample, IAB-BAMF-SOEP Survey of Refugees and SOEP Innovation Sample with administrative data of the Institute for Employment Research (IAB).

Keywords

German administrative micro data, labour market data, data manual, German Socio-Economic Panel, SOEP, survey data, linked employer-employee data, record linkage, migration data, refugees

Danksagung

Für ihre Unterstützung danken wir den Kolleginnen und Kollegen des FDZ im IAB. Für den vorliegenden Datenreport wurden Passagen aus IAB-internen Datendokumentationen der Abteilung Daten- und IT-Management (DIM) übernommen. Wir danken dem ifas Institut für angewandte Sozialwissenschaften für die Unterstützung. Wir danken Herbert Brücker, Jan Eckert, Tanja Fendel, Tobias Graf, Stephan Griesemer, Markus Köhler, Yuliya Kosyakova, Claudia Lehnert, Amrei Maddox, Andreas Moczall, Wenke Niehues, Martina Oertel, Nina Rother, Andreas Schneider, Silvia Schwanhäuser, Ehsan Vallizadeh. Außerdem danken wir allen an SOEP-Core v40.1 und SOEP IS 2023 beteiligten Kolleginnen und Kollegen am DIW Berlin.

Datenverfügbarkeit

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

1 Einleitung und Kurzbeschreibung

1.1 Einleitung

Dieser Datenreport beschreibt das Datenprodukt „Befragungsdaten der SOEP-Core-, IAB-SOEP Migrationsstichprobe, IAB-BAMF-SOEP Befragung von Geflüchteten und SOEP-Innovationssample verknüpft mit administrativen Daten des IAB“ (SOEP-CMI-ADIAB 7523), das gemeinsam vom Sozio-oekonomischen Panel (SOEP) am Deutschen Institut für Wirtschaftsforschung (DIW Berlin) und dem Institut für Arbeitsmarkt- und Berufsforschung (IAB) angeboten wird.¹ Die verknüpften Daten bestehen aus Befragungsdaten des Sozio-oekonomischen Panels (SOEP) 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 Befragungsdaten in SOEP-CMI-ADIAB umfassen die Daten von SOEP-Core, einschließlich der IAB-SOEP Migrationsstichprobe (IAB-SOEP-MIG, Brücker et al., 2024) und der IAB-BAMF-SOEP Befragung von Geflüchteten (IAB-BAMF-SOEP, Brücker et al., 2023), bis einschließlich Welle 40 (SOEP-Core v40.1eu, Goebel et al., 2025) sowie das SOEP-Innovationssample (SOEP-IS 2023, Cornesse et al., 2025; vgl. Richter und Schupp, 2015) der Jahre 1998-2023. Während IAB-SOEP-MIG und IAB-BAMF-SOEP konzeptionell Teilstichproben von SOEP-Core sind, ist das SOEP-IS eine gesonderte Panelstudie. Der Übersichtlichkeit halber werden all diese Befragungsdaten im Folgenden zusammenfassend als SOEP-CMI bezeichnet.

Das SOEP ist eine seit 1984 im jährlichen Abstand durchgeführte Längsschnitterhebung von Personen in Privathaushalten in Deutschland. Institutionell ist das SOEP am DIW Berlin angesiedelt und wird unter dem Dach der Leibniz-Gemeinschaft vom Bundesministerium für Bildung und Forschung (BMBF) und den Ländern gefördert. Weitere Hintergrundinformationen finden sich auf der Website des SOEP (siehe <https://www.diw.de/de/soep>).

Das SOEP ist eine Mehrthemenbefragung die Informationen aus unterschiedlichsten Lebensbereichen abdeckt. Hierzu gehören Aspekte wie Demographie, Einkommen, Arbeitsmarkt, Gesundheit, Zufriedenheiten, Sorgen, aber psychologische Konstrukte wie das Big-5 Inventory zur Beschreibung der Persönlichkeitsstruktur. Das SOEP verfolgt ein Haushaltskonzept, d. h. es werden alle (erwachsenen) Personen in einem Haushalt eingeladen an der Befragung teilzunehmen. In den letzten Jahren wurden zudem altersspezifische Befragungsinstrumente für Kinder und Jugendliche eingeführt, um die besondere Lebenssituation zu diesem Lebensalter zu erfragen. Neben sich jährlich wiederholenden Befragungsinhalten werden Schwerpunktthemen in mehrjährigen Abständen regelmäßig wiederholt. Einen solchen thematischen Schwerpunkt bilden Fragen zum privaten Vermögen, die seit 2002 in fünf-jährigen Abstand wiederholt werden.

Bereits im Jahr 1984 verfolgte das SOEP das Ziel eigenständige Analysen über Migrantinnen und Migranten zu unterstützen. Dazu wurde bereits in der ersten Erhebungswelle eine Überrepräsentation von Migranten in der Stichprobenziehung vorgenommen. Im Juni 1990 und damit vor der Währungs- und Wirtschaftsunion wurde das SOEP auf das Gebiet der DDR ausgeweitet. Seitdem gab es wiederholt Auffrischungstichproben um die Panelattrition

¹ Im Rahmen eines Kooperationsprojekts mit dem DIW Berlin und dem IAB leistete das Forschungsinstitut zur Zukunft der Arbeit GmbH (IZA) in den Jahren 2019 – 2022 erhebliche konzeptionelle und finanzielle Unterstützung. So wurde ein Teil der Kosten für die Erhebung der Verknüpfungserlaubnis für einige der Stichproben von SOEP-CMI übernommen.

auszugleichen als auch die zwischenzeitlichen Migrationsbewegungen nach Deutschland nachzuvollziehen. Hierzu gehören auch die IAB-SOEP Migrationsstichproben sowie die IAB-BAMF-SOEP Stichproben von Geflüchteten. Diese unterschiedlichen Teilsamples von Migranten sind im Standarddatenprodukt des Sozio-oekonomischen Panels (SOEP-Core) integriert.

Zielgruppe der IAB-SOEP Migrationsstichprobe sind Personen mit Migrationshintergrund im Sinne der Definition des Statistischen Bundesamtes, also Personen, die selbst zugewandert sind, die eine ausländische Staatsangehörigkeit haben, oder bei denen mindestens ein Elternteil zugewandert ist oder eine ausländische Staatsangehörigkeit besitzt. Bei der Durchführung des Projektes kooperiert das IAB mit dem Sozio-oekonomischen Panel (SOEP) am DIW Berlin. Beide Forschungseinrichtungen bringen in gleichem Umfang Mittel für die Erhebung der Stichprobe ein. Der Teil des IAB wurde zur Hälfte vom Bundesministerium für Arbeit und Soziales (BMAS) finanziert. Die befragten Personen wurden aus dem Kreis aller Individuen ausgewählt, die erstmals nach 1994 in den integrierten Erwerbsbiografien (IEB) des IAB auftauchen. Für die Stichprobenziehung erfolgte zunächst eine Gruppierung verfügbarer Adressen von ca. 2.500 Personen in geografische Stichproben-Regionen. Als Zielpersonen galten zum einen diejenigen, die in den Daten eindeutig als ausländisch identifiziert werden, bspw. aufgrund einer nicht-deutschen Staatsangehörigkeit, und auch diejenigen, die an spezifischen Maßnahmen für Migranten teilgenommen haben. Zum anderen wurden für die Identifikation von Personen mit Migrationshintergrund onomastische Verfahren (sprachliche Analysen von Personennamen) angewendet. Die Stichprobe für die Befragung wurde schließlich auf Grundlage soziodemografischer Merkmale aus den IEB des IAB gezogen und die erste Welle wurde 2013 durchgeführt.

Die IAB-BAMF-SOEP Befragung von Geflüchteten ist eine jährliche Wiederholungsbefragung von Schutzsuchenden und ihren Haushaltsmitgliedern, die erstmals im Jahr 2016 durchgeführt wurde. Die Stichprobe der befragten Personen wurde aus dem Ausländerzentralregister (AZR) gezogen und ist repräsentativ für die Asylbewerber, die zwischen dem 01.01.2013 und 31.12.2016 nach Deutschland eingereist sind und bis spätestens Ende Juni 2016 einen Asylantrag gestellt haben. Durch eine Aufstockung der Befragung in 2017 werden auch Schutzsuchende, die bis zum 31.12.2016 zugezogen und bis zum 01.01.2017 registriert waren, berücksichtigt. Asylbewerber mit höherer Bleibewahrscheinlichkeit zum Zeitpunkt der Stichprobenziehung wurden aufgrund des längsschnittlichen Designs bei der Ziehung überdurchschnittlich berücksichtigt. Zusätzlich wurden überdurchschnittlich viele Frauen und Personen über 30 Jahren gezogen. Bei der Durchführung des Projektes kooperiert das IAB mit dem Forschungszentrum Migration, Integration und Asyl des Bundesamtes für Migration und Flüchtlinge (BAMF-FZ) sowie mit dem Sozio-oekonomischen Panel (SOEP) am DIW Berlin. Die erste Welle der Befragung fand im Jahr 2016 statt. Durch eine Aufstockung der Befragung im Jahr 2017 werden auch Schutzsuchende berücksichtigt, die bis zum 31.12.2016 zugezogen und bis zum 01.01.2017 registriert waren. Eine weitere Aufstockung erfolgte im Jahr 2020, wodurch auch Schutzsuchende berücksichtigt werden, welche zwischen Januar 2017 und Juni 2019 zugezogen sind.

Die Durchführung der Interviews erfolgt durch geschulte Interviewerinnen und Interviewer des Instituts KANTAR Public (ehemals TNS Infratest Sozialforschung) sowie seit 2021 durch das infas Institut für angewandte Sozialwissenschaft GmbH. Die Befragung selbst wird primär

interviewerbasiert per Laptop (CAPI) oder per PAPI durchgeführt. Seit wenigen Jahren wird zudem eine webbasierte Befragung (CAWI) angeboten. Proxy-Interviews sind im SOEP nicht zulässig.

Das Standarddatenprodukt wird als SOEP-Core bezeichnet und beinhaltet sowohl die IAB-SOEP Migrationsstichproben als auch die IAB-BAMF-SOEP Stichproben von Geflüchteten. Daneben gibt es seit 2012 ein weiteres Datenprodukt, das SOEP-Innovationssample (SOEP-IS), welches ebenfalls Teil von SOEP-CMI-ADIAB ist. Das Ziel von SOEP-IS ist eine Dateninfrastruktur für WissenschaftlerInnen zur Verfügung zu stellen, um innovative Erhebungskonzepte zu testen und Experimente im Rahmen einer Panelstichprobe durchzuführen. Einen Überblick über die vielseitigen innovativen Module findet sich im [SOEP-IS Companion](#). Das SOEP-IS umfasst Personen in Privathaushalten. Es liegen Informationen von Personen ab dem Erhebungsjahr 1998 vor. Zudem wurden seit 2009 Zusatzstichproben ausschließlich für das SOEP-IS gezogen.

Die administrativen Daten stammen aus den Integrierten Erwerbsbiografien (IEB) des IAB. Sie enthalten im vorliegenden Fall Daten zu allen Personen, die im Rahmen von SOEP-CMI befragt wurden, ihre Einwilligung für die Verknüpfung erteilt 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 die Teilnahme an arbeitsmarktpolitischen Maßnahmen in der Maßnahmeteilnahmehistorik (MTH und XMTH) erfasst wird.

Neben diesen personenbezogenen Daten können auch Angaben zu den beschäftigenden Betrieben aus dem Betriebs-Historik-Panel (BHP) des IAB (Ganzer et al., 2024) 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

SOEP-CMI-ADIAB 7523 enthält schwach anonymisierte Daten, die 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 kontrollierter Datenfernverarbeitung ausgewertet werden können.

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 durch das IAB-FDZ muss, sofern noch nicht geschehen, ein Datenweitergabevertrag für die SOEP-Daten mit dem Forschungsdatenzentrum des Sozio-oekonomischen Panels (FDZ SOEP, https://www.diw.de/de/diw_01.c.678568.de/forschungsdatenzentrum_soep.html) geschlossen werden. Sobald für das beantragte Projekt ein Datenweitergabevertrag mit dem DIW Berlin besteht, wird ein Nutzungsvertrag zwischen der Institution der Forscherin bzw. des Forschers und dem IAB geschlossen. Details zur Beantragung des verknüpften Datensatzes und den Datenverarbeitungsmöglichkeiten können der Website des IAB-FDZ entnommen werden. Der Datenzugriff beginnt somit erst, wenn für das beantragte Projekt Verträge mit beiden Forschungsdatenzentren bestehen.

Zur Beantwortung von Fragen zu SOEP-CMI-ADIAB können sich Nutzerinnen und Nutzer an die Mitarbeiterinnen und Mitarbeiter des IAB-FDZ und des FDZ SOEP wenden. Die Arbeitsteilung der beiden Einrichtungen sieht vor, dass sich die Nutzerinnen und Nutzer mit Fragen zu den SOEP-Daten ausschließlich an das FDZ SOEP 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 SOEP kann neben dem [SOEPcompanion](#) auch das FDZ SOEP und dort insbesondere die SOEPHotline (soepmail@diw.de) konsultiert werden.

1.2.2 Datenhaltung

Die Befragungsdaten des SOEP 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 dem SOEP entstammende Personen-ID (pid) zusammenführen, die auch in den administrativen Personendaten enthalten ist.

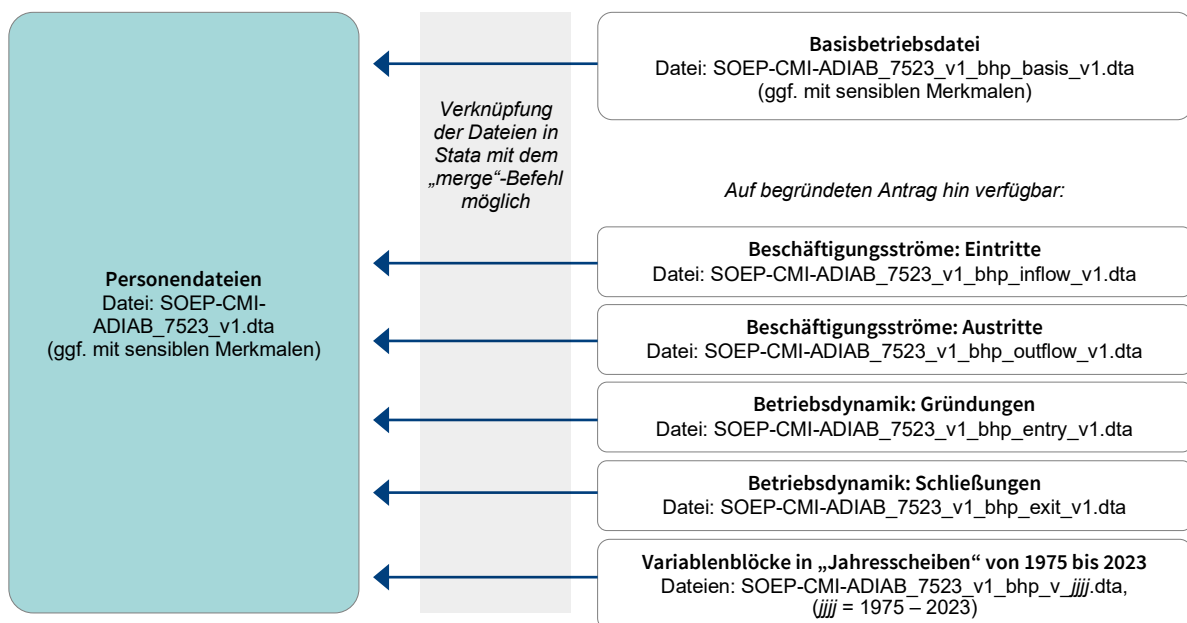
Ein weiteres Element von SOEP-CMI-ADIAB ist die Verknüpfungsdatei, welche die Personnummer sowie technische Merkmale darüber beinhaltet, ob bzw. wie die Verknüpfung mit den administrativen Daten gelungen ist. Die Datei umfasst auch die Befragten des SOEP, zu denen keine administrativen Daten gefunden werden konnten oder die keine Einwilligung zur Verknüpfung erteilt haben. Somit können Ausfallanalysen auf Basis aller Teilnehmerinnen und Teilnehmer des SOEP durchgeführt werden.

Die administrativen Daten, die analog zu den SOEP-Daten sowohl deutsche als auch englische Labels enthalten, 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 2023. Ein zweites Modul, die Basisbetriebsdatei, enthält die Betriebsnummer, das Jahr, Angaben zu Arbeitsort und Wirtschaftszweigen sowie weitere Betriebsinformationen zum Stichtag 30. Juni. Für Ostdeutschland liegen die Betriebsdaten erst ab 1992 vor.

Somit sind Betriebs- und Personenmerkmale organisatorisch voneinander getrennt, was die Struktur der Daten deutlich macht und Speicherplatz spart (vgl. Abbildung 1). Die Verknüpfung von Basisbetriebs- und Personendatei erfolgt über die Betriebsnummer und das Jahr des Datensatzes. Zu beachten ist, dass das für die Verknüpfung beider Module notwendige Merkmal „Betriebsnummer“ sowohl in der Personen- als auch in der Betriebsdatei vorhanden ist. Alle Dateinamen in der SOEP-CMI-ADIAB 7523 sind ebenfalls Abbildung 1 zu entnehmen.

Abbildung 1 Datenhaltung der administrativen Daten



1.2.3 Sensible Merkmale und Zusatzvariablen

Bestimmte, sogenannte sensible Merkmale in den administrativen Daten, die eine Deanonymisierung von Personen oder Betrieben erleichtern könnten, werden nur dann im Original weitergegeben, wenn es für das Analyseziel notwendig ist und dies im Antrag auf den Datenzugang explizit begründet wird. Es ist zudem 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. Bei der Bereitstellung eines Datensatzes mit sensiblen Merkmalen werden die entsprechenden nicht-sensiblen Merkmale gelöscht, wenn diese von Nutzenden selbst wieder generiert werden können, z. B. Bundesland- aus Kreisinformationen.

Diese aus datenschutzrechtlicher Sicht besonders sensiblen Merkmale sind:

Administrative Personendatei:

- Staatsangehörigkeit (nation)
- Geburtsmonat (gebmon)
- Berufsuntergruppe (beruf2010_4)
- Maßnahmeart – Gruppe (mass)
- Wohnort Arbeitsagentur (wo_aa)
- Wohnort Kreis (wo_kreis)

Basisbetriebsdatei (einzelne Merkmale):

- Arbeitsort Kreis (ao_kreis)
- Wirtschaftszweig – Wirtschaftsunterklasse (w93_5)
- Wirtschaftszweig – Wirtschaftsunterklasse (w03_5)
- Wirtschaftszweig – Wirtschaftsunterklasse (w08_5)

Auf begründeten Antrag können aus dem Betriebs-Historik-Panel (BHP) weitere Betriebsmerkmale in Variablenblöcken bereitgestellt werden (vgl. Abbildung 1). Zudem können die Erweiterungsmodule Beschäftigungsströme (Eintritte/Austritte) und Betriebsdynamik (Gründungen/Schließungen) mit gesonderter Begründung beantragt werden.

Diese Zusatzblöcke bzw. -module sind:

BHP-Variablenblöcke in „Jahresscheiben“:

- Struktur der Beschäftigten allgemein (az_f az_reg az_azubi az_atz az_tz az_f_vz az_f_tz az_reg_vz)
- Struktur der Beschäftigten nach Schul- und Berufsbildung (az_gq az_mq az_hq az_gq_vz az_mq_vz az_hq_vz)
- Altersstruktur der Beschäftigten (az_15_19 az_15_19_vz az_20_24 az_20_24_vz az_25_29 az_25_29_vz az_30_34 az_30_34_vz az_35_39 az_35_39_vz az_40_44 az_40_44_vz az_45_49 az_45_49_vz az_50_54 az_50_54_vz az_55_59 az_55_59_vz az_60_64 az_60_64_vz az_ab65 az_ab65_vz alter_mw alter_mw_vz)
- Aktivitäten Forschung und Entwicklung (az_ingnat)
- Anzahl atypisch Beschäftigter (az_leih az_bfr)
- Struktur der Beschäftigten nach Staatsangehörigkeit (az_d az_d_vz az_eu)
- Struktur der Beschäftigten nach Blossfeld Berufsqualifikation (az_bf_agr az_bf_emb az_bf_edi az_bf_evb az_bf_qmb az_bf_qdi az_bf_qvb az_bf_tec az_bf_semi az_bf_ing az_bf_prof az_bf_man)
- Struktur der Beschäftigten nach Tätigkeitsniveau (az_niv1 az_niv2 az_niv3 az_niv4)
- Gehaltsstruktur der Vollzeitbeschäftigten (az_zens te_imp_med te_imp_p25 te_imp_p75 te_imp_mw_f te_imp_med_f te_imp_med_m te_imp_mw_gq te_imp_med_gq te_imp_mw_mq te_imp_med_mq te_imp_mw_hq te_imp_med_hq te_imp_med_uq te_imp_mw_d te_imp_med_d te_imp_med_a te_imp_sum)

BHP-Erweiterungsmodule:

- Beschäftigtenströme (ein_ges ein_f ein_reg ein_gf ein_azubi ein_vz ein_tz ein_f_vz ein_f_tz ein_reg_vz ein_wdr ein_wdr_f ein_bw ein_bw_f ein_bf_agr ein_bf_emb ein_bf_edi ein_bf_evb ein_bf_qmb ein_bf_qdi ein_bf_qvb ein_bf_tec ein_bf_semi ein_bf_ing ein_bf_prof ein_bf_man ein_15_19 ein_20_24 ein_25_29 ein_30_34 ein_35_39 ein_40_44 ein_45_49 ein_50_54 ein_55_59 ein_60_64 ein_ab65 aus_ges aus_f aus_reg aus_gf aus_azubi aus_vz aus_tz aus_f_vz aus_f_tz aus_reg_vz aus_temp aus_bw aus_bw_f aus_bf_agr aus_bf_emb aus_bf_edi aus_bf_evb aus_bf_qmb aus_bf_qdi aus_bf_qvb aus_bf_tec aus_bf_semi aus_bf_ing aus_bf_prof aus_bf_man aus_15_19 aus_20_24 aus_25_29 aus_30_34 aus_35_39 aus_40_44 aus_45_49 aus_50_54 aus_55_59 aus_60_64 aus_ab65 aus_senio_1 aus_senio_2 aus_senio_3)
- Betriebsdynamik (eintritt besch besch_vor status_vor inflow austritt besch besch_nach status_nach outflow)

Nähere Informationen zum BHP finden sich unter <https://fdz.iab.de/betriebsdaten/betriebs-historik-panel-bhp-version-7523-v1> und im FDZ-Datenreport 09/2024 (Ganzer et al., 2024).

Zudem stehen personen- und betriebsspezifische Effekte (AKM-Effekte) für die Jahre 1985 bis 2023 zur Verfügung (Dateinamen: SOEP-CMI-ADIAB_7523_v1_akm_estab.dta; SOEP-CMI-ADIAB_7523_v1_akm_pers.dta). Weitere Informationen sind im FDZ-Methodenreport 03/2025 (Lochner/Wolter, 2025) zu finden. Bei Neuanträgen können die AKM-Effekte über das Feld für die sensiblen Merkmale und Ergänzungsmodule beantragt werden. Bei bestehenden Projekten reicht eine formlose E-Mail an iab.fdz@iab.de.

Die Daten des SOEP-IS können im Rahmen von SOEP-CMI-ADIAB ebenfalls genutzt werden. Sie müssen jedoch als Erweiterungsmodul zusätzlich beantragt und ihre Notwendigkeit für das Forschungsprojekt gesondert begründet werden. Die Nutzung kleinteiliger Regionalinformationen, wie bspw. Gemeindeganziffern ist im Rahmen der Nutzung von SOEP-CMI-ADIAB nicht möglich.

Für die Nutzung von SOEP-CMI-ADIAB im Rahmen von Gastaufenthalten an unseren Standorten außerhalb der EU müssen die Daten aus datenschutzrechtlichen Gründen zusätzlichen Anonymisierungsmaßnahmen unterzogen wurden. Eine der Maßnahmen besteht darin, dass nur die internationale Version der Befragungsdaten bereitgestellt wird, in der die Zahl der Personen um 5% reduziert wurde. Die davon betroffenen Fälle werden auch aus den in SOEP-CMI-ADIAB enthaltenen administrativen Personen- und Betriebsdaten entfernt. Weitere Anonymisierungsmaßnahmen werden projektspezifisch und in Abstimmung mit dem IAB-FDZ vorgenommen.

1.3 Änderungen zur Vorgängerversion

1.3.1 Änderungen in den Befragungsdaten

Die in SOEP-CMI-ADIAB 7523 verfügbare SOEP-Version ist die SOEP v40.1. Die Neuerungen in SOEP v40.1 sind gesammelt auf der Website des FDZ SOEP unter dem Reiter „SOEP-Core v40.1 (2023) - Änderungen am Datensatz“ zu finden. Da die Version SOEP v40.1 eine Aktualisierung von SOEP v40 ist, sei ebenfalls auf die Änderungen am Datensatz v40 verwiesen, die an gleicher Stelle zu finden sind.

SOEP-Core v40.1 wurde um eine zusätzliche Variable ergänzt, die es Nutzerinnen und Nutzern mit einem gültigen Datenweitergabevertrag mit dem SOEP ermöglicht, bereits vor Beantragung von SOEP-CMI-ADIAB 7523 zu überprüfen, wie viele und welche SOEP-Befragten Teil von SOEP-CMI-ADIAB 7523 sind. Die Variable *reclin_iab_match* findet sich im SOEP Datensatz *ppathl*. Eine Person hat in der Variable die Ausprägung [1], wenn sie erfolgreich mit den administrativen Daten verknüpft werden konnte und dementsprechend Teil von SOEP-CMI-ADIAB 7523 ist. Nutzerinnen und Nutzer können sich so frühzeitig ein Bild über die Größe der für sie relevanten Stichprobe machen und daraus Schlüsse über die Realisierbarkeit ihres Projekts ziehen.

1.3.2 Merkmalskanon

In SOEP-CMI-ADIAB 7523 wurden gegenüber SOEP-CMI-ADIAB 7521 neue Merkmale zum Tagesentgelt aufgenommen, die bei der Anrechnung von Einmalzahlungen unterstützen und eine Imputation der zensierten Werte rechts der Beitragsbemessungsgrenze anbieten.² Die zugrundeliegenden Datenarbeiten werden in Drechsler et al. (2023) bzw. Drechsler/Ludsteck (2025) beschrieben. Weitere Informationen zu den neuen Merkmalen finden sich unter „Tagesentgelt (inkl. Einmalzahlungen) (*tentgelt_bonus*)“ und „Tagesentgelt (imputiert) (*tentgelt_imp*)“.

Des Weiteren ist das sensible Merkmal „Maßnahmeart – Gruppe (*mass*)“ neu enthalten. Dieses ist fachlich den quellspezifischen Angaben zu MTH/XMTH des Merkmals „Erwerbsstatus (*erwstat*)“ zugeordnet und bietet detailliertere Informationen zu spezifischen Maßnahmeteilnahmen. Das Merkmal ist nur auf begründeten Antrag hin erhältlich.

Beim Merkmal „Wohnort – Arbeitsagentur (*wo_aa*)“ gibt es Änderungen bei den Werten und Wertelabels aufgrund der Gebietsreformen in Sachsen-Anhalt und Thüringen. Die Angaben für die Jahre vor 1999 wurden allerdings nicht gebietsstandsbereinigt und erhalten deshalb einen neuen Wertebereich. Weitere Informationen finden sich im Online-Anhang (siehe Abschnitt 8.1).

Tabelle 1 gibt einen Überblick über die Änderungen im Merkmalskanon über alle Quellen hinweg.

Tabelle 1 Neue, entfallene und veränderte Merkmale in SOEP-CMI-ADIAB 7523

Merkmal	Erläuterung
Tagesentgelt (inkl. Einmalzahlungen) (<i>tentgelt_bonus</i>)	Neues Merkmal
Tagesentgelt (imputiert) (<i>tentgelt_imp</i>)	Neues Merkmal
Maßnahmeart - Gruppe (<i>mass</i>)	Neues Merkmal, detaillierte Maßnahmeart-Differenzierung des IAB
Wohnort -Arbeitsagentur (<i>wo_aa</i>)	Neue Werte und Wertelabels aufgrund der Gebietsreformen in Sachsen-Anhalt und Thüringen
Erstes Auftreten Betriebsnummer (<i>grd_dat</i>)	Wird nicht länger angeboten; <i>grd_jahr</i> weiter vorhanden

² Das in den Skripten von Stüber et al. (2023) enthaltene Verfahren zur Imputation der Löhne kann nun durch die neu angebotenen Variablen ersetzt werden.

Merkmal	Erläuterung
Letztes Auftreten Betriebsnummer (lzt_dat)	Wird nicht länger angeboten; lzt_jahr weiter vorhanden

1.3.3 Datendokumentation

Im vorliegenden Datenreport wurden die Beschreibungen für folgende Merkmale überarbeitet bzw. ergänzt:

- Ausbildung (ausbildung)
- Tagesentgelt/ täglicher Leistungssatz (tentgelt)
- Tagesentgelt (inkl. Einmalzahlungen) (tentgelt_bonus)
- Tagesentgelt (imputiert) (tentgelt_imp)
- Maßnahmeart – Gruppe (mass)
- Restanspruch/geplante Dauer (restanspruch)

Der Abschnitt „Merkmalsliste“ wurde gelöscht. Als Ersatz dienen die detaillierten Übersichten zu den einzelnen Merkmalen (Label, Auszählungen, fehlende Werte) im Online-Anhang zum Datenreport (siehe Abschnitt 8.1).

1.4 Steckbrief

Tabelle 2 Steckbrief SOEP-CMI-ADIAB

Kategorie	Beschreibung
Themen / Merkmalsgruppen	<p><u>Befragungsdaten</u> Längsschnittdaten zu Haushaltszusammensetzung; Persönlichkeitsmerkmalen; Soziodemographischen Charakteristika; Arbeitsmarktbeteiligung und Einkommen; Gesundheit; (Lebens-)Zufriedenheit; Herkunft und Weg nach Deutschland; Stand des Asylverfahrens; Unterkunft in Deutschland; Materielle Ausstattung und gesellschaftliche Teilhabe; Bekanntheit, Bedarf und Nutzung von Unterstützung- und Beratungsangeboten Sprachkenntnisse und Sprachkursbesuch; Schul- Hochschul- und Berufsbildung im Ausland und in Deutschland; Soziale Netzwerke.</p> <p><u>Administrative Personendaten</u></p> <p>Beschäftigtenhistorik (BeH): Jahres- und Abmeldungen zur Sozialversicherung für sozialversicherungspflichtig und geringfügig Beschäftigte</p> <p>Leistungsempfängerhistorik (LeH): Informationen zu Leistungsbezug nach SGB III für Personen, die Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld beziehen</p> <p>Leistungshistorik Grundsicherung (LHG): 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>Arbeitsuchendenhistorik (ASU):</p>

Kategorie	Beschreibung
	<p>Informationen zur Arbeitsuche, die in BA-Verfahren erfasst werden</p> <p>Arbeitssuchendenhistorik aus XSozial-BA-SGB II (XASU): Informationen zur Arbeitsuche, die über den Übermittlungsstandard XSozial-BA-SGB II von zugelassenen kommunalen Trägern an die BA gemeldet werden.</p> <p>Maßnahmeteilnahmehistorik (MTH): Informationen zu Teilnahmen an Maßnahmen der aktiven Arbeitsmarktförderung (keine Maßnahmen der zugelassenen kommunalen Träger im SGB II)</p> <p>Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II (XMTH): 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>Kerndatensatz: 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>Erweiterungsmodul-Beschäftigtenströme: 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>Erweiterungsmodul-Betriebsdynamik: Gründungsjahr, Art der Gründung, Hilfsvariablen zur Klassifizierung der Gründung, Schließungsjahr, Art der Schließung, Hilfsvariablen zur Klassifizierung der Schließung</p>
Untersuchungseinheit	<p><u>Befragungsdaten</u> Zielpopulation sind Personen in Privathaushalten in Deutschland. Eine Übersicht zu den Teilstichproben des SOEP befindet sich im SOEPcompanion.</p> <p><u>Administrative Personendaten</u> Sozialversicherungspflichtig Beschäftigte (ab 1999 auch geringfügig Beschäftigte), Personen im Leistungsbezug, Arbeitssuchende, Maßnahmeteilnehmende, Beschäftigungsbetriebe</p>
Fallzahlen	<p><u>Administrative Daten</u></p> <p>44.162 Personen</p> <p>1.395.095 Originalsätze</p> <p>2.215.505 überschneidungsfreie Sätze (nach Episodensplitting)</p> <p>Von den 44.162 Personen, zu denen administrative Daten vorliegen, entstammen</p> <ul style="list-style-type: none"> • 41.758 Personen aus SOEP-Core; darunter: <ul style="list-style-type: none"> ○ SOEP-Core (ohne IAB-SOEP-MIG/IAB-BAMF-SOEP): 21.699 ○ IAB-SOEP-MIG: 7.442 ○ IAB-BAMF-SOEP: 12.617 • 2.404 Personen aus SOEP-IS

Kategorie	Beschreibung
	<p>In IAB-SOEP-MIG und IAB-BAMF-SOEP wurde teilweise bereits seit 2013 nach Einwilligung zur Verknüpfung gefragt. In den restlichen Stichproben von SOEP-Core sowie in SOEP-IS wurde erstmals 2019 nach Einwilligung zur Verknüpfung gefragt. Durch die unterschiedlichen Zeitpunkte der Einwilligung zur Verknüpfung sowie die unterschiedlichen Zeitpunkte von Eintritten und Austritten in die und aus der SOEP-Befragung, liegen Befragungsdaten für verschiedene Personen zu unterschiedlichen Jahren vor.</p>
Zeitraum	<p><u>Befragungsdaten</u> Seit 1984 jährlich, Schwerpunkt der Befragungen im 1. Halbjahr.</p> <p><u>Administrative Daten</u> Der Zeitraum hängt von der Datenquelle ab.</p> <p>BeH: 01.01.1975 - 31.12.2023 (bis 2020: 36-Monats-Datei, 2021 und 2022: 18-Monats-Datei, 2023: 6-Monats-Datei)</p> <p>LeH: 01.01.1975 - 31.12.2023</p> <p>ASU: 01.01.1997 - 31.12.2023</p> <p>LHG: 01.01.2005 - 31.12.2023</p> <p>XASU: 01.01.2005 - 31.12.2023</p> <p>MTH: 01.01.2000 - 31.12.2023</p> <p>XMTH: 01.01.2005 - 31.12.2023</p>
Zeitbezug	<p><u>Befragungsdaten</u> In der Regel Befragungszeitpunkt; Einkommens- und Tätigkeitskalendarium retrospektiv, teilweise bezogen auf Vorjahr, Vormonat, oder Zeitraum seit letzter Befragung. Vereinzelt prospektiv bezogen auf künftige Erwartungen oder Vorhaben.</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 31.12.2023</p>
Erhebungsdesign	<p><u>Befragungsdaten</u> Jährlich wiederholendes Befragungsprogramm, dazu Schwerpunktmodule mit unterschiedlichen zeitlichen Abständen, wie z.B. das Vermögensmodul, welches alle 5 Jahre erhoben wird.</p> <p>Der Großteil der Befragung wird als CAPI (computer-assisted personal interviewing) durchgeführt, kleinere Anteile entfallen auf persönliche Interviews, selbstständig ausgefüllte Fragebögen sowie auf Web Interviews. Die genauen Anteile der durchgeführten Interviews in den jeweiligen Interviewmodi können dem SOEP Methodenbericht (Weitz et al. 2025) entnommen werden.</p> <p><u>Administrative Daten</u> Identifikation der Befragten mit erteilter Verknüpfungserlaubnis in den Integrierten Erwerbsbiografien des IAB</p>

Kategorie	Beschreibung
Beteiligte Institutionen	<p><u>Befragungsdaten</u></p> <p>Auftraggeber: Sozio-oekonomisches Panel am DIW Berlin. Finanziert durch das Bundesministerium für Forschung, Technologie und Raumfahrt und dem Land Berlin. Durchführung der Befragung: 1984-2020: Kantar (ehemals TNS Infratest) Ab 2021: infas Institut für angewandte Sozialwissenschaft GmbH</p> <p><u>Administrative Daten</u></p> <p>Sozialversicherungsträger, Bundesagentur für Arbeit, kommunale Träger</p>
Frequenz der Datensammlung	<p><u>Befragungsdaten</u></p> <p>Jährlich</p> <p><u>Administrative Daten</u></p> <p>Laufend</p>
Dateiformat/ -größe	<p><u>Befragungsdaten</u></p> <p>SOEP-Core (Stata): 590 Einzeldatensätze zwischen 29 KB und 7,1 GB SOEP-IS (Stata): 31 Einzeldatensätze zwischen 78 KB und 374 MB</p> <p><u>Administrative Daten</u></p> <p>Stata; Personendatensatz: 188,3 MB; Betriebsdatensatz: 29,4 MB</p>
Dateiorganisation	<p><u>Befragungsdaten</u></p> <p>Die Daten sind in mehreren Dateien abgelegt, die im Quer- oder Längsschnitt (wellen- bzw. episodenspezifisch) angelegte Personen- und Haushaltsdaten enthalten. Auf begründeten Antrag werden zusätzlich die Daten des SOEP-IS bereitgestellt.</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 Verknüpfungserfolg beinhaltet.</p>
Datenzugang	Gastaufenthalt am IAB-FDZ, kontrollierte Datenfernverarbeitung beim IAB-FDZ nach vorangegangenem Gastaufenthalt
Anonymisierungsgrad	<p><u>Befragungsdaten</u></p> <p>Ausschließlich anonymisiert</p> <p><u>Administrative Daten</u></p> <p>Schwach anonymisiert</p>
Sensible Merkmale	<p><u>Administrative Daten</u></p> <p>Geburtsmonat (gebmon), Staatsangehörigkeit (nation), Maßnahmeart – Gruppe (mass), Berufsuntergruppe (beruf2010_4), Wohnort – Kreis (wo_kreis), Wohnort – Arbeitsagentur (wo_aa), Arbeitsort – Kreis (ao_kreis), Wirtschaftsunterklasse (w93_5), Wirtschaftsunterklasse (w03_5), Wirtschaftsunterklasse (w08_5)</p>
Zitierung der Daten und Datendokumentation	<p><u>Daten</u></p> <p>Antoni, Manfred; Beckmannshagen, Mattis; Grabka, Markus M.; Keita, Sekou; Trübswetter, Parvati; Creators DOI:10.5684/soep.core.v40.1eu; Creators DOI:10.5684/soep.iab-bamf-soep-mig.2023.1; Creators DOI:10.5684/soep.iab-soep-mig.2023.1; Creators DOI:10.5684/soep.is.2023; Daten- und IT-Management (Team Datenproduktion); Deutsches Institut für Wirtschaftsforschung (DIW Berlin) (2026):</p>

Kategorie	Beschreibung
	<p>„Befragungsdaten der SOEP-Core-, IAB-SOEP Migrationsstichprobe, IAB-BAMF-SOEP Befragung von Geflüchteten und SOEP-Innovationsstichprobe verknüpft mit administrativen Daten des IAB (SOEP-CMI-ADIAB) – Version 7523 v1“. Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB). DOI: 10.5164/IAB.SOEP-CMI-ADIAB7523.de.en.v1</p> <p>Der Datenzugang erfolgt über einen Gastaufenthalt am Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ) und anschließend mittels kontrollierter Datenfernverarbeitung beim FDZ.</p> <p><u>Datendokumentation</u></p> <p>Antoni, Manfred; Beckmannshagen, Mattis; Grabka, Markus M.; Keita, Sekou; Trübswetter, Parvati (2026): Befragungsdaten der SOEP-Core-, IAB-SOEP Migrationsstichprobe, IAB-BAMF-SOEP Befragung von Geflüchteten und SOEP-Innovationsstichprobe verknüpft mit administrativen Daten des IAB (SOEP-CMI-ADIAB) 1975-2023. FDZ-Datenreport, 02/2026(de), Nürnberg. DOI: 10.5164/IAB.FDZD.2602.de.v1</p> <p>SOEP-Core: Goebel, Jan; Grabka, Markus M.; Liebig, Stefan; Kroh, Martin; Richter, David; Schröder, Carsten; Schupp, Jürgen (2019): The German Socio-Economic Panel (SOEP). Jahrbücher für Nationalökonomie und Statistik 239 (2), 345-360. DOI:10.1515/jbnst-2018-0022</p> <p>IAB-SOEP-MIG: Brücker, Herbert; Kroh, Martin; Bartsch, Simone; Goebel, Jan; Kühne, Simon; Liebau, Elisabeth; Trübswetter, Parvati; Tucci, Ingrid; Schupp, Jürgen (2014): The new IAB-SOEP Migration Sample: an introduction into the methodology and the contents. SOEP Survey Paper 216, Series C. Berlin, Nürnberg: DIW Berlin. DOI: 10.5684/soep.iab-soep-mig.2020.</p> <p>IAB-BAMF-SOEP: Brücker, Herbert; Rother, Nina; Schupp, Jürgen (2017): IAB-BAMF-SOEP Befragung von Geflüchteten 2016. Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Qualifikation, Sprachkenntnissen sowie kognitiven Potenzialen. IAB Forschungsbericht 13/2017 (de), Nürnberg.</p> <p>SOEP-IS: Richter, David; Schupp, Jürgen (2015). The SOEP Innovation Sample (SOEP IS). Schmollers Jahrbuch 135 (3), 389-399. DOI: 10.3790/schm.135.3.389</p>
Datensatzversion	Befragungsdaten der SOEP-Core-, IAB-SOEP Migrationsstichprobe, IAB-BAMF-SOEP Befragung von Geflüchteten und SOEP-Innovationsstichprobe verknüpft mit administrativen Daten des IAB (SOEP-CMI-ADIAB 7523); DOI: 10.5164/IAB.SOEP-CMI-ADIAB7523.de.en.v1

1.5 Mengengerüst

In Tabelle 3 ist die Anzahl der Episoden nach Quelle sowie vor und nach dem Episodensplitting, welches in Abschnitt 3.2 näher erläutert wird, abgebildet.

Tabelle 3 Mengengerüst

Fallzahlen	vor Episodensplitting	nach Episodensplitting
BeH	819.323	1.049.857
LeH	80.367	135.25

Fallzahlen	vor Episodensplitting	nach Episodensplitting
LHG	129.568	349.183
ASU	262.859	489.183
XASU	46.724	86.46
MTH	44.698	79.604
XMTH	11.556	25.968
Sätze insgesamt	1.395.095	2.215.505
Personen		44.162

2 Datenquellen

2.1 Das Sozio-oekonomische Panel (SOEP)

Das Sozio-oekonomische Panel bildet eine Datengrundlage für Forschung zu einer Vielzahl von Themen innerhalb zahlreicher wissenschaftlicher Disziplinen. Durch die Panelstruktur und das Befragen von Personen und Haushalten über Generationen hinweg, lassen sich Informationen über das gesellschaftliche Zusammenleben in der Bundesrepublik Deutschland ableiten, die sonst nicht verfügbar wären. Die verschiedenen im SOEP erhobenen Fragemodule lassen sich dabei in 11 Themenbereiche gliedern:

- Demographie und Bevölkerung
- Arbeit und Beschäftigung
- Einkommen, Steuern, soziale Sicherung
- Familie und soziale Netzwerke
- Behausung, Haushaltsorganisation und -ausgaben
- Bildung und Qualifizierung
- Einstellung, Werte, Persönlichkeitsmerkmale
- Zeitverwendung und Umweltbewusstsein
- Migration und Integration
- Surveymethodik

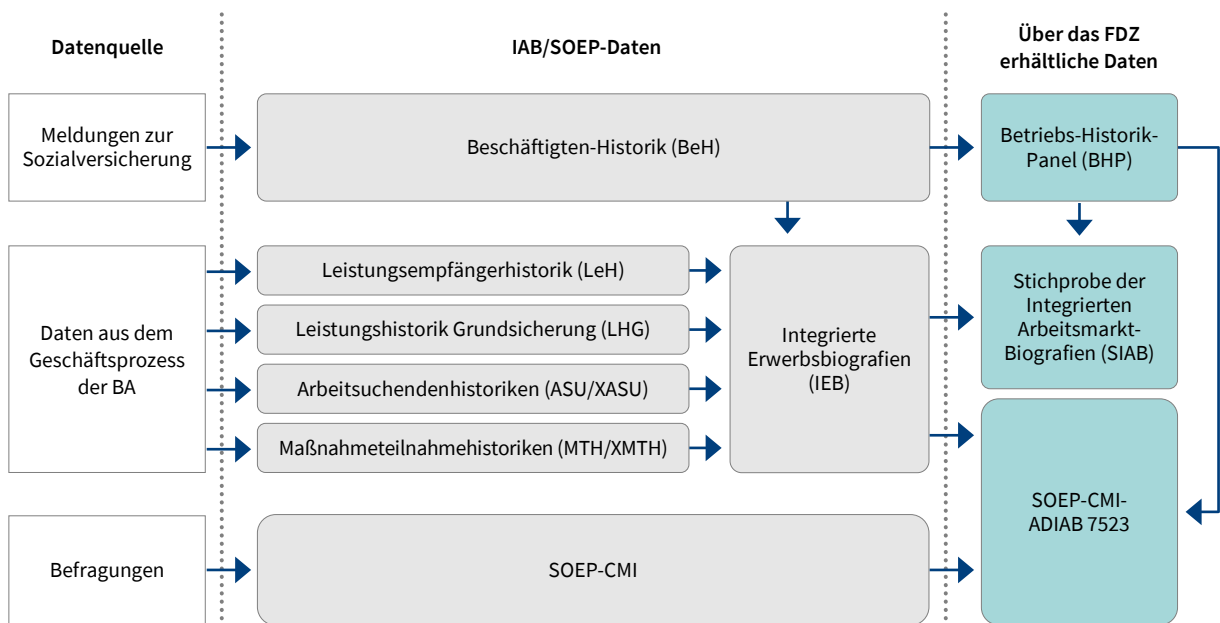
Weitere Informationen zu den Themenbereichen sowie einzelne Beschreibungen der Fragemodule können dem [SOEPcompanion](#) entnommen werden. Die IAB-SOEP Migrationsstichprobe liefert zusätzlich Detailinformationen zu Personen mit Migrationshintergrund in Deutschland und ihre Haushaltsmitglieder. Sie beinhaltet Informationen zu deren familiärer Situation, sozialer Herkunft, Migrations- und Integrationserfahrung, Persönlichkeitsmerkmalen und Wertvorstellungen. Die IAB-BAMF-SOEP-Befragung von Geflüchteten umfasst die Migrations-, Bildungs- und Erwerbsbiografien von Geflüchteten, die Fluchtursachen und -wege und erhebt die Persönlichkeitsmerkmale, Einstellungen und Informationen zum Gesundheitszustand der Geflüchteten. Andere wichtige Teile der Studie sind die Unterbringung, Asylverfahren, Integration in den deutschen Arbeitsmarkt, das Bildungssystem und andere gesellschaftliche Bereiche sowie begleitende Politikmaßnahmen.

Darüber hinaus umfasst das SOEP-IS eine Vielzahl an innovativen Erhebungsmodulen, die sich entweder methodisch oder inhaltlich nicht in die SOEP-Core Befragung integrieren lassen. Die Inhalte des SOEP-IS ergeben sich primär aus einem Ausschreibungs- und Reviewverfahren, wobei u.a. nach Innovativität und Publikationsaussichten der Ergebnisse selektiert wird. Einen Überblick über die existierenden Erhebungsmodule des SOEP-IS kann der projektspezifischen [Homepage](#) sowie dem eigens für das SOEP-IS existierende [SOEP-IS Companion](#) entnommen werden. Da das SOEP-IS eine unabhängig vom SOEP-Core gezogene Stichprobe ist, beinhaltet es auch eigene Gewichtungsfaktoren. Eine gemeinsame Nutzung von SOEP-IS und SOEP-Core wird nicht unterstützt. So liegen z. B. keine gemeinsamen Gewichtungsfaktoren für beide Studien vor.

2.2 Administrative Daten aus den Integrierten Erwerbsbiografien (IEB)

Die administrativen Personendaten werden aus den Integrierten Erwerbsbiografien (IEB, Institut für Arbeitsmarkt- und Berufsforschung, 2024) des IAB gezogen und hinsichtlich Variablenauswahl und -ausprägungen entsprechend der vom FDZ erstellten Stichprobe der integrierten Arbeitsmarktbiografien (SIAB, siehe Schmucker et al., 2025) 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 SOEP-CMI-ADIAB führen und stellt deren Beziehung zu anderen Datenprodukten des IAB-FDZ dar.

Abbildung 2 Quellen von SOEP-CMI-ADIAB 7523



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 an die zuständigen Sozialversicherungsträger abzugeben. Die BeH erfasst alle Arbeiter*innen 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. Studierende können dennoch in der BeH erscheinen, wenn sie parallel zum Studium z. B. geringfügig beschäftigt sind. 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. Die Basisbetriebsdaten und BHP-Erweiterungsmodule, die im Rahmen von SOEP-CMI-ADIAB bereitgestellt werden, umfassen nur Angaben zu Betrieben, für die mindestens eine beschäftigte Person in den administrativen Personendaten von SOEP-CMI-ADIAB enthalten ist. Des Weiteren sind die Betriebsdaten auf die Jahre begrenzt, in denen die beschäftigten Personen in den administrativen Personendaten vorzufinden sind.

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. Dies bedeutet

im Extremfall, dass für einen Betrieb keine Informationen zu einem Jahr im BHP vorliegen, wenn zu diesem Betrieb aufgrund von Konkurs nur Beschäftigungsmeldungen vom 01.01. bis zum 31.05. des betreffenden Jahres vorliegen.

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

Kasten 1 Beispielcode zur Verbindung von Personen- und Betriebsdaten in Stata 18

```
use $orig/SOEP-CMI-ADIAB_7523_v1.dta, clear
gen int jahr = year(begepi)
sort betnr_soepadiab jahr
merge m:1 betnr_soepadiab jahr using $orig/SOEP-CMI-
ADIAB_7523_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 und von der Bundesagentur für Arbeit übernommene Beiträge zur privaten Kranken- bzw. Pflegeversicherung während des Leistungsbezugs. Darüber hinaus sind auch Informationen über Sperr- und Unterbrechungszeiten enthalten:

- Eine Sperrzeit ist ein Zeitraum, in dem eine prinzipiell leistungsberechtigte Person aus Gründen, die sie selbst zu vertreten hat, keine Leistung erhält. Typisch ist z.B. eine kurze Sperrzeit zu Beginn eines Leistungsanspruchs, weil die Person sich verspätet arbeitsuchend gemeldet hat.
- Eine Unterbrechungszeit ist hingegen ein Zeitraum ohne Leistungserhalt, den die leistungsberechtigte Person nicht selbst zu vertreten hat. Das kann z.B. eine zwingende Ortsabwesenheit sein. Eine Unterbrechungszeit verringert nicht die Anspruchsdauer, d. h. die Person erhält zwar während der Unterbrechungszeit keine Leistung, dafür verschiebt sich aber das reguläre Ende des Leistungsanspruchs entsprechend nach hinten.

Leistungen nach dem Rechtskreis SGB II (z.B. Arbeitslosengeld II) sind nicht enthalten. 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. Daher enthält auch SOEP-CMI-ADIAB keine Angaben zur Höhe des SGB-II-Leistungsbezugs.

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

- Arbeitsgemeinschaften (ARGE, bis 2010) / gemeinsame Einrichtungen (ab 2011) aus gemeinsamer Aufgabenwahrnehmung von BA und Kommune,
- Getrennte Trägerschaft (gT)/getrennte Aufgabenwahrnehmung (bis 2011) – hier besteht eine Aufgabenteilung zwischen BA und Kommune³,
- Zugelassene kommunale Träger, die nach der anfänglichen Experimentierklausel des § 6a auch als optierende Kommunen oder Optionskommunen bezeichnet werden – hier übernimmt die Kommune alle Aufgaben aus dem Bereich des SGB II.

Die Daten der LHG stammen aus verschiedenen Meldeverfahren. In der Regel findet das IT-Verfahren A2LL bis 2010 bei ARGE, ab 2011 bei gemeinsamen Einrichtungen Verwendung.⁴ Zugelassene kommunale Träger verwenden verschiedene eigene IT-Verfahren und übermitteln ihre Daten mittels des Standards X Sozial-BA-SGB II einmal monatlich an die Bundesagentur für Arbeit. Bei Kommunen mit getrennter Trägerschaft kommen beide Verfahren zum Einsatz. Die unterschiedlichen Datenstandards haben Auswirkungen auf Umfang und Qualität der gelieferten Daten.

Der linke Datenrand der LHG ist der 01.01.2005. Die Datenquelle ist jedoch bis Anfang 2007 lückenhaft (siehe Kapitel 5).

2.6 Arbeitsuchendenhistoriken (ASU und XASU)

Daten über Arbeitsuchende werden in den Arbeitsuchendenhistoriken (ASU und XASU) gespeichert. Die Quelle ASU enthält Informationen zu Arbeitsuchenden, welche bei Arbeitsagenturen verzeichnet sind, ab 2005 auch solche, die bei Arbeitsgemeinschaften und

³ 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.

⁴ 2014 wurde A2LL sukzessive durch ALLEGRO als neues IT-Verfahren für die Grundsicherung im SGB II in gemeinsamen Einrichtungen ersetzt.

getrennten Trägerschaften gemeldet sind. In der Quelle XASU sind ab 2005 die Daten von arbeitsuchenden 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.

3 Datenaufbereitung und Stichprobenziehung

3.1 Bereinigungen und Aufbereitungen

3.1.1 SOEP-Befragungsdaten

Detaillierte Informationen zur Stichprobenziehung, zur Datenerhebung, zu den eingesetzten Instrumenten sowie zur Struktur und Aufbereitung der SOEP Daten finden sich auf der SOEP-Website (siehe <https://www.diw.de/soep>). Das SOEP liefert neben den direkt aus den einzelnen Fragen hervorgehenden Variablen mehrere generierte Datensätze, die die Nutzung der SOEP Daten für die Forschungscommunity erleichtern sollen. Diese generierten Datensätze werden Plausibilitätschecks und Qualitätskontrollen unterzogen. Einen Überblick über die generierten Datensätze sowie einen Verweis auf jeweilige detaillierte Dokumentationen findet sich im [SOEPcompanion](#).

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 z.B. das 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 ausgeschlossen, deren Daten nicht für den gesamten Beobachtungszeitraum zur Verfügung stehen. 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 ab 2011 neu eingeführt.
- Aus datenschutzrechtlichen Gründen werden die Personengruppen 107, 111, 113, 114, 127 und 204 zur Personengruppe "Sonstige Erwerbstätige" (599) zusammengefasst.
- Ab dem Meldejahr 2012 wurden Auszubildende mit den Personengruppen 121 und 122 neu aufgenommen.
- Sätze mit 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 31.12.2023.

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.
- Eine Gebietsstandsberreinigung wird analog zur BeH durchgeführt.

3.1.5 Leistungshistorik Grundsicherung (LHG)

- Sätze ohne BA-Kundennummer werden gelöscht.
- Stornierte Datensätze werden nicht genutzt.
- Es sind nur Sätze erwerbsfähiger Personen und von Personen über 65 Jahren enthalten.

- Es sind jeweils überschneidungsfreie Zeiträume der Leistungsberechtigung einer Person in einer bestimmten Bedarfsgemeinschaft abgebildet. Aus folgenden administrativen Gründen beginnt ein neuer Satz:
 - an gesetzlich festgelegten, für strukturelle Änderungen der Bedarfsgemeinschaft relevanten Geburtstagen (14, 15, 18 und 25) sowie dem individuell gültigen Renteneintrittsalter der Mitglieder einer Bedarfsgemeinschaft (siehe Abschnitt 3.1.9),
 - bei Veränderungen in der Zusammensetzung einer Bedarfsgemeinschaft (z.B. durch Zu-/Abgänge),
 - bei Merkmalsänderungen des Bedarfsgemeinschafts-Kunden sowie
 - bei Beginn und Ende eines Sanktionssachverhalts für Sätze ab dem 01.01.2006. Allerdings ist zu beachten, dass anhand der Daten weder die Dauer, die Art noch der Zeitpunkt der Verhängung bzw. des Beginns der Sanktion festgestellt werden kann. Grund hierfür ist das Fehlen eines entsprechenden Merkmals oder einer Ausprägung, welches den Beginn, die Art oder die Dauer der Sanktion anzeigt.
- Aus obigem Grund sind alle personenbezogenen Merkmale, die für die Quelle LHG gefüllt sind, für den gesamten Zeitraum des Satzes gültig.
- Doppelmeldungen aufgrund der Gebietsreformen 2009/2011 und Trägerneuorganisation 2012 (siehe Abschnitt 5.2.2) wurden soweit möglich bereinigt.
- Eine Gebietsstandsberreinigung wird analog zur BeH durchgeführt.

3.1.6 Arbeitssuchendehistoriken (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 (X)ASU 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 Datensatz gebildet:
 - Wechsel des Arbeitssuche-Status
 - Wechsel der Verfügbarkeit
 - Wechsel des SGB-II-Trägers
 - Wechsel des Wohnortes
- Eine Gebietsstandsberreinigung wird analog zur BeH durchgeführt.

3.1.7 Maßnahmeteilnehmehistoriken (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 Datensatz zu einer einzelnen Maßnahmeteilnahme verwendet.
- In die IEB werden nur Maßnahmeteilnahmen aufgenommen, die als „tatsächlich stattgefunden“ klassifiziert sind. Teilnahmen, die nicht oder noch nicht stattgefunden haben, werden gelöscht. Teilnahmen werden auch dann als nicht stattgefunden gewertet, wenn während der Teilnahme an einer Maßnahme ein Löschdatum gesetzt wird.
- Bestimmte Maßnahmearten werden nicht aufgenommen. Hierzu zählen Unterstützung der Beratung und Vermittlung, Mobilitätshilfen und reine Reha-Maßnahmen.
- Eine Gebietsstandsberreinigung wird analog zur BeH durchgeführt.

3.1.8 Maßnahmeteilnehmehistoriken 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 und 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 teilnehmender Person 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 Maßnahmeart (aus XSozial) werden zu einer Beobachtung zusammengefasst. Maßnahmespezifische 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 Gebietsstandsberreinigung wird analog zur BeH durchgeführt.

3.1.9 SGB-II-Anonymisierung

Um das Deanonymisierungsrisiko zu verringern, wird in den administrativen Daten in SOEP-CMI-ADIAB standardmäßig nur das Geburtsjahr ausgewiesen. Der Geburtsmonat kann als sensibles Merkmal nur bei begründetem Bedarf beantragt werden. 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 an diesem Tag neu beginnen.

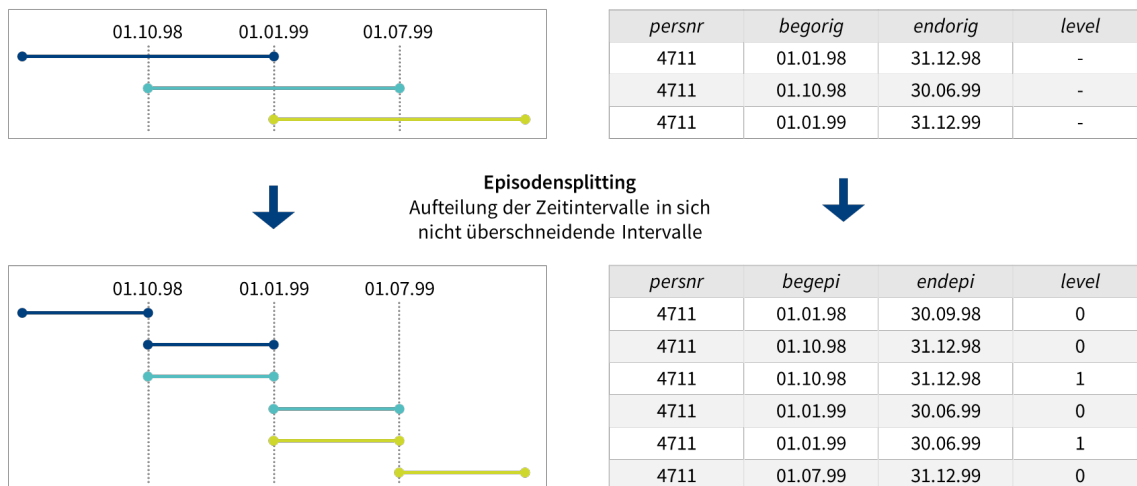
Um eine indirekte Bestimmung des exakten Geburtstags zu verhindern, wird daher das folgende Verfahren angewendet. Sätze, die am 18., 25. oder 65. Geburtstag geteilt wurden, werden zu einem einzelnen Satz zusammengefügt. Hierdurch geht außer dem exakten Geburtstag keine weitere Information verloren. Bei 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 endepi) kennzeichnen Anfang und Ende der gesplitteten Episoden. Ob Sätze gesplittet wurden, kann an einem Vergleich des ursprünglichen Zeitraums (begorig und endorig) mit dem Episodenzeitraum (begepi und endepi) festgestellt werden. Um die Originaldaten ohne Episodensplitting 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 18

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

3.3 Fehlende Werte

In den Daten gilt die in folgende Codierung von fehlenden Werten:

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

4 Datenverknüpfung

4.1 Informierte Zustimmung der Befragten zur Datenverknüpfung

Ausgangspunkt der Verknüpfung der Erhebungs- und administrativen Daten sind die Befragten von SOEP-CMI. Diese wurden im Rahmen des Interviews um Erlaubnis gebeten, ihre Befragungsdaten für Forschungszwecke mit den über sie am IAB vorhandenen administrativen Daten verknüpfen zu dürfen. Dabei wurden die Befragte wie folgt informiert und nach Zustimmung gefragt (hier beispielhaft die Computer-basierte Abfrage aus dem Erhebungsjahr 2023):

Leben in Deutschland

Wir würden bei der Auswertung der Befragung gerne Auszüge aus Daten einbeziehen, die beim Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit in Nürnberg über Sie 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?

- Ja
- Nein
- Kann / möchte ich nicht beantworten

Tabelle 4 gibt einen Überblick darüber, wie viele Befragte von SOEP-CMI bis einschließlich 2023 nach Zustimmung zur Datenverknüpfung gefragt wurden und wie viele diese erteilt haben. Wenn Personen im Zeitverlauf mehrfach um ihr Einverständnis zur Verknüpfung gebeten wurden und unterschiedliche Antworten (Zustimmung/Verweigerung) gegeben wurden, wurde als Grundlage für das vorliegende Datenprodukt die spätere bzw. jeweils letzte Antwort herangezogen. Es werden also nur Personen in Betracht gezogen, die der Frage nach Einverständnis zur Verknüpfung zugestimmt haben, ohne dass diese Zustimmung zu einem späteren Zeitpunkt zurückgezogen wurde. Die durchschnittliche Zustimmungsrates zur Datenverknüpfung über alle Bestandteile von SOEP-CMI lag bei 75,3%.

Tabelle 4 Abfrage und Erteilung der Einwilligung zur Datenverknüpfung in SOEP-CMI

	SOEP-Core (ohne Mig und Ref)	SOEP Mig (Teilstichpro- ben M1, M2, M7, M8)	SOEP Ref (Teilstichpro- ben M3, M4, M5, M6, M9)	SOEP-IS	SOEP-CMI (Summe)
Nach Einwilligung gefragt	30.810	12.420	17.454	4.280	64.964
Einwilligung gegeben (bis einschl. 2023)	23.461	8.302	14.509	2.623	48.895
Anteil	76,1%	66,8%	83,1%	61,3%	75,3%

Hinweis: Die Berechnung erfolgt nur auf Grundlage von Personen mit vollständigem Personeninterview. Code zur Replikation dieser Tabelle ist auf Nachfrage verfügbar. Aufgrund eines nachträglichen Widerrufs der Einwilligung zur Verknüpfung weichen die Zahlen leicht von denen aus den Variablen in SOEP-Core v40.1 ab. Befragte des Bonner Interventions-Panels im SOEP-IS wurden für SOEP-CMI-ADIAB 7523 nicht berücksichtigt.

4.2 Herkunft der Verknüpfungsmerkmale

Für die Verknüpfung der Befragungsdaten der im SOEP 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 SOEP-Erhebung wurden diese Merkmale vom jeweils zum Zeitpunkt der Verknüpfung zuständigen Befragungsinstitut 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, wenn sie nicht schon zuvor verknüpft waren, bei einer neuen Verknüpfung nicht berücksichtigt. Unabhängig von einer ggf. bereits erfolgten Verknüpfung werden Personen, die zwischenzeitlich eine Löschung ihrer Befragungsdaten gefordert haben, nicht in die verknüpften Daten berücksichtigt. Dadurch besteht eine Differenz zwischen der Anzahl der Personen, die im Rahmen der SOEP-Befragung der Verknüpfung irgendwann zugestimmt haben und der Anzahl der Personen, die für das Linkage zur Verfügung standen.

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 Geburtsjahrgänge und Postleitzahl-Bereiche der zuspielbereiten Befragten 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 Verknüpfungsmerkmale der Erhebungs- und administrativen Daten wurden Bereinigungsroutinen 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) am IAB-FDZ für die Besonderheiten des deutschen Kontexts verfeinert wurden (siehe Antoni/Schnell, 2019 oder <https://www.record-linkage.de>). 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 Piktographen entfernt. Mehrere Vor- bzw. Nachnamen wurden jeweils in einem Feld ohne

Leerzeichen zusammengefasst. Neben diesen grundlegenden Bereinigungen wurden variablenspezifische 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 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. 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 fälschlicherweise zugeordnet werden könnten. 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 sukzessive Abweichungen bei jeweils 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 dennoch 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 ToolBox (siehe Schnell et al., 2004) durchgeführt. Für Vor- und Nachname sowie Ort und Straße 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 eine iterative Blockingstrategie angewandt. In einem ersten Schritt wurde ein Blocking der Beobachtungen auf Ebene der fünfstelligen 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 fünfstelliger Postleitzahl 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 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 -monat sowie jeweils auf einen phonetischen Code des Vornamens (NYSIIS, siehe Borgman/Siegfried 1992), dreistellige Postleitzahl und Geburtsjahr durchgeführt (siehe Tabelle 5).

Tabelle 5 Iterative Blockingstrategie

Schritt	Blockingvariable(n)
1	Postleitzahl (5-Steller)
2	Geschlecht, Geburtsjahr und -monat
3	NYSIIS-Code des Vornamens
4	Postleitzahl (3-Steller)
5	Geburtsjahr

Bei der Berechnung des Ähnlichkeitsmaßes wurde ein probabilistisches Linkage angewandt, das 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 wurde pro befragter Person jeweils die Entsprechung in den administrativen Daten mit dem höchsten Ähnlichkeitswert ausgewählt und manuell überprüft. Dabei wurden nur plausible Matches behalten und unplausible Matches entfernt. Uneindeutige Fälle wurden als potentielle Matches markiert, um sie später einer genaueren Prüfung zu unterziehen. Für die verbliebenen Personen, denen kein passender Match zugewiesen werden konnte, wurde dieses Vorgehen im darauffolgenden Blockingschritt wiederholt.

Zuletzt wurde für alle Personen mit einem potentiellen Match, für die durch keinen der o. g. Schritte ein passender Match gefunden werden konnte, ein manuelles Linkage durchgeführt. Hierzu wurden alle Einträge zu den in den bisherigen Schritten gesammelten potentiellen Matches in beiden Datenquellen gesichtet und plausible Matches ausgewählt. Aufgrund der manuellen Prüfungen war die Wahrscheinlichkeit äußerst gering, im Rahmen des fehlertoleranten Linkage unterschiedliche Personen als Match zu klassifizieren.

4.6 Ergebnis der Datenverknüpfung

Insgesamt 48.895 Personen erteilten bis einschließlich 2023 ihre Einwilligung zur Verknüpfung (75,3% aller um Einwilligung gebetenen Befragten). Für 45.152 Personen konnten die Befragungsdaten einem Konto in den administrativen Daten zugeordnet werden. Dies entspricht einer sehr hohen Verknüpfungsquote von 92% der Personen, die für eine Zuspiegelung der administrativen Daten zur Verfügung standen. Es verblieben lediglich 3.761 Personen, die trotz Verknüpfungserlaubnis nicht in den administrativen Datenbeständen gefunden wurden. Hinzu kommen 990 Personen, für die es einen Match in den administrativen Adressdatenbeständen gibt, für die jedoch keine IEB-Sätze existieren. Weitere Erläuterungen, weshalb für einzelne Personen keine IEB-Sätze existieren, befinden sich in Abschnitt 5.2. In Tabelle 6 werden die Erfolgsquoten der einzelnen Verknüpfungsschritte zusammengefasst.

Generierte technische Merkmale im Datensatz (siehe Abschnitt 6.8) kennzeichnen für jede befragte Person, mit welcher Verknüpfungsmethode 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 6 Ergebnisse des Linkage-Prozesses

Personengruppe	N	Anteil (in %)
Befragte mit Verknüpfungseinwilligung	48.895	100,0
Erfolgreich verknüpfte Befragte	45.152	92,3
Erfolgreich verknüpfte Befragte mit IEB-Sätzen	44.162	90,3
Darunter:		
Exaktes Matching	38.745	79,2
Probabilistisches Matching	5.343	10,9
Manuelles Matching	74	0,2

5 Datenqualität und Probleme

5.1 Sozio-oekonomisches Panel (SOEP)

Eine präzise Beschreibung der Daten, deren Struktur sowie Informationen zur Stichprobe und zur Befragung allgemein finden sich im [SOEPcompanion](#) und den dort verlinkten Dokumentationen. Bei Fragen und Problemen kann außerdem die [SOEPHotline](#) kontaktiert werden.

5.2 Gesamte IEB

5.2.1 Lücken in den Erwerbsbiografien

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 Abmeldegrund / Abgabegrund / Beendigungsgrund und Erwerbsstatus in den verschiedenen Quellen. Die Auflistung in Tabelle 7 erhebt keinen Anspruch auf Vollständigkeit.

Tabelle 7 Biographische Lücken und Identifikationsmöglichkeiten

Biografische Lücke	Information über Lücke, ggf. identifizierbar anhand der Ausprägung des Merkmals „grund“ im Vorgängersatz der Quelle
Beamte*innen, Berufssoldat*innen, Richter*innen, Angestellte von Körperschaften oder Stiftungen des öffentlichen Rechts	XASU
Ungeförderte Selbstständige	LeH, ASU
Studierende, Personen in schulischer Weiterbildung	LeH, LHG, ASU, XASU
Personen, die länger als 6 Wochen krank / arbeitsunfähig sind (Krankheit während Arbeitslosigkeit ist 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
Beziehende von Vorruhestandsgeld	LeH, ASU
Hausgewerbetreibende	
Kurzfristig Beschäftigte	ASU
Personen in Einrichtungen der Jugendhilfe, Berufsbildungswerken, anerkannten Werkstätten oder ähnlichen Einrichtungen für Menschen mit Behinderung	ASU
Teilnehmende an Leistungen zur Teilhabe am Arbeitsleben (Rehabilitanden)	ASU
(Nebenerwerbs-)Landwirt*innen	
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

Biografische Lücke	Information über Lücke, ggf. identifizierbar anhand der Ausprägung des Merkmals „grund“ im Vorgängersatz der Quelle
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-Beziehende (vor Einführung des SGB II 2005), Sozialgeld-Beziehende (im Sinne des SGB II)	
SGB-II-Leistungsberechtigte, bei deren Trägern Lieferausfälle vorkamen	
Ausgleichsempfangende nach dem FELEG (Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit)	

5.2.2 Einführung des SGB II und spätere Trägerwechsel

Mit der Einführung des SGB II zum 01.01.2005 wurden auch die Zuständigkeiten für die Betreuung von Arbeitssuchenden neugestaltet, sodass die Aufgabenwahrnehmung im SGB II von unterschiedlichen Trägertypen übernommen werden kann:

- In der Regel übernahmen die Arbeitsgemeinschaften (ARGE) aus BA und Kreis die Aufgaben im Bereich des SGB II. Diese wurde 2011 von den gemeinsamen Einrichtungen (gE) – auch „Jobcenter“ genannt - abgelöst.
- Zudem war es bis Ende 2011 möglich, dass BA und Kreis die die ihnen übertragenen Aufgaben in eigener (getrennter) Zuständigkeit – in sogenannten getrennten Trägerschaften (gT) oder auch getrennten Aufgabenwahrnehmungen (gAw) erbrachten.
- Seit 2005 besteht zudem die Möglichkeit, dass zugelassene kommunale Träger (zKT; auch: optierende Kommunen) die Aufgaben übernehmen. Zunächst erhielten 69 Landkreise oder kreisfreie Städte die alleinige Trägerschaft der Grundsicherung für Arbeitsuchende bis 31.12.2010. Mit der Verfassungsänderung wurde die Optionsmöglichkeit ab 2012 auf 110 Träger ausgedehnt.

Während die ARGEN/gE das Leistungs- und Fallmanagement über die BA-Verfahren pflegen und die erfassten Daten an die Statistik der BA übermitteln, verwenden die kommunalen Grundsicherungsträger jeweils eigene Softwaresysteme. Die Übermittlung an die Statistik der BA erfolgt über den Standard XSozial. Für getrennte Träger lag eine Aufteilung der Zuständigkeiten vor. Der jeweilige Trägerteil war nur für die Erfassung und Lieferung der Daten verantwortlich, die in seinen Zuständigkeitsbereich fallen. Daher gab es reduzierte Datenanforderungen für die kommunale Seite der getrennten Träger. Die Übermittlung über den Standard XSozial erfolgt einmal monatlich zu einem festgelegten Zeitfenster. In der Woche darauf besteht an einem zweiten Meldetag die Möglichkeit, missglückte Lieferungen zu wiederholen. Im Lauf der Jahre wurden zahlreiche Qualitätssicherungsinstrumente entwickelt. Dennoch erhöhen

Versionswechsel des Standards oder der vor Ort genutzten Software regelmäßig das Risiko für Datenqualitätsprobleme.

Im Rahmen der Neuorganisation von SGB-II-Trägern jeweils zum 01.01. der Jahre 2011 bis 2014 gingen diverse ARGEn/gE und getrennte Träger in kommunale Trägerschaft über bzw. wurden von zkt in gE umgewandelt. Diese Trägerwechsel und der damit einhergehende Wechsel der vor Ort verwendeten Software führen teilweise auch zu Brüchen in den Daten der SGB-II-Quellen. Detaillierte Informationen zu den daraus resultierenden Qualitätsproblemen in den Daten finden sich in den einzelnen Abschnitten zu den Datenquellen.

5.3 Beschäftigtenhistorik (BeH)

- Die Beschäftigungshistorik enthält in wenigen Fällen (kleiner 1%) Duplikate in den Merkmalen Personenummer, Betriebsnummer, Erwerbsstatus, Beginndatum Originalzeitraum und Enddatum Originalzeitraum. Das Meldeverfahren sieht solche Duplikate eigentlich nicht vor. In den meisten Fällen unterscheiden sich die einzelnen Sätze der Duplikatengruppe im Tagesentgelt, in selteneren Fällen auch in anderen Merkmalen. Der Grund für solche Duplikate sind vermutlich nicht korrekt verarbeitete Stornomeldungen.
- Die Übermittlung von Informationen zur Ausbildung, ausgeübten Tätigkeit und Stellung im Beruf erfolgt durch Meldungen der Arbeitgeber nach DEÜV (siehe Abschnitt 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 Angaben 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.⁵
- Das Merkmal „Stellung im Beruf und Arbeitszeit“ (stib) ist nur für Meldungen gefüllt, die vor Einführung des neuen Tätigkeitsschlüssels abgegeben wurden. 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

⁵ Siehe Paulus/Matthes, 2013, für Details zur Klassifikation der Berufe 2010.

Sozialversicherung eingeräumt wurde,⁶ zu einer temporären Zunahme fehlender Angaben. Die im Rahmen des Meldeverfahrens eingesetzten Prüfprogramme ließen bis einschließlich Mai 2012 leere Angaben beim Tätigkeitsschlüssel 2010 zu. 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 SOEP-CMI-ADIAB 7523 verwendet. Für die Lücken in den anderen Merkmalen werden keine Imputationen vorgenommen.

- Aufgrund der Einführung des Meldeverfahrens in den neuen Bundesländern kann erst ab 1993 von einer hinreichenden Vollständigkeit der Meldungen ostdeutscher Betriebe ausgegangen werden. Aus demselben Grund weisen sehr viele Sätze des Jahres 1991 für einige Merkmale (u. a. Ausbildung, Erwerbsstatus und Tagesentgelt) fehlende Werte auf.
- Der Anstieg der Zahl der BeH-Sätze ab 1999 folgt aus der Einführung der Meldepflicht für geringfügig Beschäftigte ab dem 01.04.1999.
- Es lässt sich insbesondere 1999 ein deutlicher Anstieg bei den Meldungen zu Teilzeitbeschäftigung feststellen. Verursacht wird dieser Befund sowohl von dem tatsächlich zu beobachtenden Anstieg der Teilzeitbeschäftigung als auch von der Tatsache, dass seit 1999 die Beschäftigtenmeldungen generell besser ausgefüllt werden.
- 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 (z. B. kann frühestens im Juli 2024 die 18-Monats-Datei für das Jahr 2022 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 2020 das letzte Jahr mit einem Füllgrad der BeH-Meldungen von 100%. Für die Jahre 2021 und 2022 wurden 18-Monats-Dateien verwendet, für 2023 eine 6-Monats-Datei. Es ist daher davon auszugehen, dass Beschäftigungsmeldungen für die Jahre 2021 bis 2022 leicht und für 2023 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.

⁶ Die im Rahmen des Meldeverfahrens eingesetzten Prüfprogramme ließen bis einschließlich Mai 2012 leere Angaben beim Tätigkeitsschlüssel 2010 zu.

- Im Jahr 1984 erfolgte eine Änderung des Meldeverfahrens. Einmalig gezahltes Bruttoarbeitsentgelt wurde ab diesem Zeitpunkt als Teil des beitragspflichtigen Jahresentgelts gemeldet, was zu einer Erhöhung des durchschnittlichen Tagesentgelts führt. Insbesondere steigt ab diesem Jahr der Anteil der Entgelte oberhalb der Beitragsbemessungsgrenze stark an (vgl. Bender et al., 1996).
- Für die Jahre 1992 bis 2000 wurden in den IEB auffällige Einbrüche und Anstiege in der Anzahl der Meldungen festgestellt. Es lassen sich v.a. Einbrüche bei folgenden zehn Kreisen beobachten: Braunschweig (03101), Wolfsburg (03103), Emden (03402), Kassel (06633), Essen (05113), Neuss (05162), Erftkreis (05362), Hersfeld-Rotenburg (06632), Miltenberg (09676) und Kempten (Allgäu) (09763). Grund hierfür sind Meldeprobleme bei einzelnen bzw. mehreren Betrieben in diesen Regionen.
- Nennenswerte Einbrüche verzeichnen auch die Kreise Salzgitter (03102) und Hoyerswerda (14264).
- Bei den Meldungen zu Vollzeitbeschäftigung in den IEB sind besonders die Kreise Main-Taunus-Kreis (06436) und Alzey-Worms (07331) erwähnenswert. Hier sind überdurchschnittliche Anstiege zu beobachten. Als Grund sind ebenfalls Meldeprobleme bei einzelnen bzw. mehreren Betrieben in diesen Regionen anzuführen.
- In den Jahren 1996 bis 1998 kommen die Werte 841-844 (Ärzt*innen und Apotheker*innen) in der Variable Beruf – ausgeübte Tätigkeit (beruf) im Vergleich zu benachbarten Jahren sehr selten vor. Die Gründe dafür sind nicht bekannt.
- In den Jahren 1975 bis 1977 gab es bisher deutlich zu viele Beschäftigte mit einem Arbeitsort in der niedersächsischen Gemeinde Lahn (Kreis Emsland). Statt im zu erwartenden zweistelligen Bereich gab es pro Jahr bis zu 90.000 Beschäftigtenmeldungen mit diesem Arbeitsort in der Grundgesamtheit der BeH. Ursächlich ist eine historische Fehlzuordnung der Beschäftigtenmeldungen aus der Stadt Lahn. Hierbei handelte es sich um einen nach kurzer Zeit wieder aufgelösten Zusammenschluss der hessischen Gemeinden Wetzlar, Gießen, Heuchelheim, Wettenberg und Lahnau (Kreise Gießen und Lahn-Dill-Kreis). Auf Basis dieser Erkenntnisse und unter der Annahme, dass es keine realen Betriebsumzüge zwischen diesen Regionen gab, wurde auf Ebene der Betriebsnummer folgende Bereinigungsregel umgesetzt: Sobald der Betriebsort in den Jahren 1975 bis 1978 von der Gemeinde Lahn auf eine der aufgeführten hessischen Gemeinden wechselt, wird die frühere Angabe Gemeinde Lahn mit der späteren korrekten Angabe überschrieben. Über diese Regel wird der Überhang deutlich reduziert und die niedersächsische Gemeinde Lahn weist danach in den kritischen Jahren zumindest nur noch unter 2.000 Beschäftigtenmeldungen in der Grundgesamtheit der BeH auf.

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.

- Durch einen internen Systemwechsel kommt es 2004 zu einem Bruch in der Erfassung von Sperr- und Säumniszeiten. Bis 01.07.2004 können Sperr- und Säumniszeiten nur durch den Abmeldegrund im vorhergehenden LeH-Satz identifiziert werden. Danach existiert ein eigener Satz mit täglichem Leistungssatz = 0 für Versagens-, Entziehens- und Ruhezeiten.

5.5 Leistungshistorik Grundsicherung (LHG)

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

5.6.1 ASU

- Die registrierten Zeiten der Arbeitsuche 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 das Merkmal Erwerbsstatus vor Arbeitsuche (estatvor) zu identifizieren (Wechsel zu einem zugelassenen kommunalen Träger).
- 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 Arbeitssuchestatus 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 Arbeitsuchendmeldungen 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 Maßnahmeteilnehmende auf verschiedene Träger aufgeteilt wurden (siehe Abschnitte 2.7 und 5.2.2).
- 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. Meldungen, die zugelassene kommunale Träger über den Standard XSozial übermitteln, finden sich in der XMTH.
- Durch die Neuorganisation von SGB-II-Trägern 2011-2014 (siehe Abschnitt 5.2.2) 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).⁷
- 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.
- Durch die Neuorganisation von SGB-II-Trägern 2011-2014 (siehe Abschnitt 5.2.2) 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).

⁷ Weitere Informationen zur Gebietsstruktur der SGB-II-Träger und relevanten Änderungen finden sich unter: <https://statistik.arbeitsagentur.de/DE/Navigations/Grundlagen/Klassifikationen/Regionale-Gliederungen/Gebietsstruktur-Traeger-Grundsicherung-Nav.html>.

- Die über monatliche Stichtagsauszählungen in der XMTH erhaltenen Zahlen bzw. Personen unterscheiden sich in mehrfacher Hinsicht von den veröffentlichten BA-Statistiken.

6 Variablenbeschreibung

Dieses Kapitel beschreibt die im Datenprodukt enthaltenen Variablen im Detail. Weitere detaillierte Übersichten zu den einzelnen Merkmalen (Label, Auszählungen, fehlende Werte) sind dem Online-Anhang zum Datenreport zu entnehmen (siehe Abschnitt 8.1).

6.1 Identifikatoren

6.1.1 Personen-ID (pid)

Kategorie	Beschreibung
Variablenlabel	Unveränderliche Personennummer
Variablenname	pid
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 pid ist sowohl in den Befragungsdaten 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_soepadiab)

Kategorie	Beschreibung
Variablenlabel	Systemfreie Betriebsnummer
Variablenname	betnr_soepadiab
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 https://www.arbeitsagentur.de/unternehmen/betriebsnummern-service/alles-wichtige). 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.

Kategorie	Beschreibung
	<p>Hat das Unternehmen mehrere Niederlassungen in einer 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 nicht derselben Wirtschaftsklasse angehören.</p> <p>Hat das Unternehmen mehrere Niederlassungen in mehreren 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 Betrieb 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 Arbeitsstätte ist eine Einheit, in der Beschäftigte tätig sind und für die nach o.a. Grundsätzen keine Betriebsnummer zu vergeben ist.</p> <p>Ein Unternehmen als Begriff fasst Betriebs- und Arbeitsstätten desselben Arbeitgebers zusammen.</p> <p>Ein Arbeitgeber 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	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.

6.2 Generierte technische Merkmale

6.2.1 Satzähler pro Konto (spell)

Kategorie	Beschreibung
Variablenlabel	Satzähler pro Konto
Variablenname	spell
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	Der Satzähler pro Konto nummeriert aufsteigend die Sätze einer Person. Das Merkmal wird beim Episodensplitting erzeugt und bezieht sich auf die gesplitteten Sätze. Mit dem Merkmal „Satzähler 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.

6.2.2 Quelle des Satzes (quelle)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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. Beispielcode mit dem „merge“-Befehl in Stata:</p> <pre>use \$orig/SOEP-CMI-ADIAB_7523_v1.dta gen int jahr = year(begepi) sort betnr_soepadiab jahr merge m:1 betnr_soepadiab jahr using \$orig/SOEP-CMI-ADIAB_7523_v1_bhp_basis_v1.dta</pre>

6.3 Gültigkeitszeitraum

6.3.1 Beginndatum Originalbeobachtung (begorig)

Kategorie	Beschreibung
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>1) BeH</p>

Kategorie	Beschreibung
	<p>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>2) LHG, ASU, XASU</p> <p>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>3) MTH, XMTH</p> <p>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)

Kategorie	Beschreibung
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 Abschnitt 3.2). Da die Meldelogik unter bestimmten Umständen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben in diesen Fällen durch das im Kapitel 3.1.9 beschriebene Anonymisierungsverfahren verändert.</p> <p>1) BeH</p> <p>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>2) LHG, ASU, XASU</p> <p>Bestimmte Änderungen führen zur Anlage eines neuen Datensatzes (siehe Abschnitte 2.5 und 3.1.6). endorig gibt das Enddatum des neuen Zeitraums an.</p> <p>3) MTH, XMTH</p> <p>In der MTH und XMTH entstehen neue Datensätze, wenn ein Trägerwechsel während der Teilnahme an einer arbeitsmarktpolitischen Maßnahme stattfindet.</p>

6.3.3 Beginndatum Episode (begepi)

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

6.3.4 Enddatum Episode (endepe)

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

6.4 Informationen zur Person

6.4.1 Geschlecht (frau)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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.

Kategorie	Beschreibung
	Um aus den Angaben zum Geburtsjahr (gebjahr) und -monat (gebmon) eine Variable im Datumsformat JJJJmM (z. B. 1984m6) zu generieren, kann in Stata 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)

Kategorie	Beschreibung
Variablenlabel	Staatsangehörigkeit
Variablenname	nation
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält als Ausprägungen die Staatenschlüssel des Statistischen Bundesamtes (Statistisches Bundesamt, 2024).
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)

Kategorie	Beschreibung
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, 2024). Eine Übersicht über diese Gruppen findet sich in der „Übersicht der Klassifikationen“ unter den zentralen Arbeitshilfen (siehe https://fdz.iab.de/anleitungen-und-hilfen/zentrale-arbeitshilfen).
Qualitätshinweis	Das Merkmal ist in der LeH erst ab 1983 gut gefüllt.

6.4.6 Familienstand (famst)

Kategorie	Beschreibung
Variablenlabel	Familienstand
Variablenname	famst
Kategorie	Persönliche Merkmale

Kategorie	Beschreibung
Herkunft	LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	<p>Das Merkmal beschreibt den Familienstand. Diese Variable ist für Beschäftigungsmeldungen nicht gefüllt.</p> <p>1) LeH In der Quelle LeH umfasst das Merkmal nur zwei Ausprägungen (0 - nicht verheiratet, 1 - verheiratet).</p> <p>2) LHG, ASU, XASU, MTH, XMTH 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>1) LeH 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>2) MTH Ab 2010 weist das Merkmal immer weniger fehlende Werte auf.</p> <p>3) LHG, XASU, XMTH Aufgrund abweichender Meldestandards in XSozial sind die Angaben aus den zkT bis Dezember 2009 als weniger valide einzustufen.</p>

6.4.7 Kinderzahl (kind)

Kategorie	Beschreibung
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. Diese Variable ist für Beschäftigungsmeldungen nicht gefüllt.</p> <p>1) LeH 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: 0 Keine Kinder 100 Ein oder mehrere Kinder</p> <p>2) ASU, MTH Hier entspricht die Ausprägung des Merkmals der Anzahl an Kindern unter 15 Jahren im Haushalt.</p> <p>3) LHG, XASU, XMTH 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>

Kategorie	Beschreibung
Qualitätshinweis	<p>1) LeH 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>2) ASU, MTH 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> <p>3) XMTH Ab 2011 weist das Merkmal immer weniger fehlende Werte auf.</p>

6.4.8 Ausbildung (ausbildung)

Kategorie	Beschreibung
Variablenlabel	Ausbildung
Variablenname	ausbildung
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	<p>Das Merkmal enthält den beruflichen Ausbildungsabschluss. Eine Einordnung der Abschlüsse in die Internationale Standardklassifikation des Bildungswesens (ISCED 2011) ist den zentralen Arbeitshilfen des FDZ (https://fdz.iab.de/anleitungen-und-hilfen/zentrale-arbeitshilfen) zu entnehmen. Zu beachten ist, dass das Merkmal je nach Quelle eine unterschiedliche Bedeutung hat.</p> <p>1) BeH 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"> 1 ohne abgeschlossene Berufsausbildung 2 betriebliche/außerbetriebliche Ausbildung 11 FH ohne nähere Spezifikation 12 Hochschulabschluss ohne nähere Spezifikation <p>Bei Meldungen nach dem neuen Tätigkeitsschlüssel (siehe Kapitel 5.2.2) 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>2) ASU, MTH 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"> 1 Ohne abgeschlossene Berufsausbildung 2 Betriebliche/außerbetriebliche Ausbildung 3 Berufsfachschule 4 Fachschule 5 Fachhochschule 6 Universität

Kategorie	Beschreibung									
	<p>Im Jahr 2006 wurde das EDV-Verfahren, aus dem die Arbeitsuchendendaten 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 Ausprägungen vorliegen:</p> <ul style="list-style-type: none"> 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 26 Duales Studium, ausbildungsintegrierend 27 Duales Studium, praxisintegrierend 28 Bachelor 29 Master/Diplom/Magister 30 Staatsexamen, kirchl. Examen <p>3) XASU, XMTH</p> <p>Für diese Sätze ist die letzte abgeschlossene Berufsausbildung enthalten. Folgende Ausprägungen kommen vor:</p> <ul style="list-style-type: none"> 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 <p>Für Auswertungen über einen längeren Zeitraum können die Ausprägungen folgendermaßen aggregiert werden:</p> <table data-bbox="542 1803 1189 1937" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%;">1, 7, 8</td> <td style="width: 30%;">auf</td> <td style="width: 20%; text-align: right;">1</td> </tr> <tr> <td>2, 3, 4, 9</td> <td>auf</td> <td style="text-align: right;">2</td> </tr> <tr> <td>5, 6, 10, 11 - 21, 23 - 30</td> <td>auf</td> <td style="text-align: right;">3</td> </tr> </table> <p>Die aggregierten Werte haben folgende Labels:</p> <ul style="list-style-type: none"> 1 Ohne (anerkannte) abgeschlossene Berufsausbildung 	1, 7, 8	auf	1	2, 3, 4, 9	auf	2	5, 6, 10, 11 - 21, 23 - 30	auf	3
1, 7, 8	auf	1								
2, 3, 4, 9	auf	2								
5, 6, 10, 11 - 21, 23 - 30	auf	3								

Kategorie	Beschreibung
	2 Betriebliche / schulische Ausbildung 3 Akademische Ausbildung
Qualitätshinweis	<p>1) BeH</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>Mit der Umstellung des Tätigkeitsschlüssels ist ein Bruch in den Daten verbunden, der über die reine Umstellung des Schlüssels hinausgeht, und evtl. durch Aktualisierungseffekte bei der Umstellung der Lohnbuchhaltungssoftware bei den meldenden Betrieben zu erklären ist.</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>2) ASU, MTH</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>3) XASU, XMTH</p> <p>Der Füllgrad in den Quellen XASU und XMTH ist generell niedrig.</p>

6.4.9 Ausbildung (imputiert) (ausbildung_imp)

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

Kategorie	Beschreibung
	<p>Das Imputationsverfahren wird in Thomsen et al. (2018) beschrieben, basierend auf der Arbeit von Fitzenberger et al. (2006). 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> <p>Eine Einordnung der Abschlüsse in die Internationale Standardklassifikation des Bildungswesens (ISCED 2011) ist den zentralen Arbeitshilfen des FDZ (https://fdz.iab.de/anleitungen-und-hilfen/zentrale-arbeitshilfen) zu entnehmen.</p>
Qualitätshinweis	Trotz des Imputationsverfahrens nimmt der Anteil der fehlenden Angaben im Zeitverlauf zu und erreicht 2019 ca. 9%. Mit der Umstellung des Tätigkeitsschlüssels ist ein Bruch in den Daten verbunden, der über die reine Umstellung des Schlüssels hinausgeht, und evtl. durch Aktualisierungseffekte bei der Umstellung der Lohnbuchhaltungssoftware bei den meldenden Betrieben zu erklären ist.

6.4.10 Schulausbildung (schule)

Kategorie	Beschreibung
Variablenlabel	Schulausbildung
Variablenname	schule
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	<p>Das Merkmal enthält den Schulabschluss. Eine Einordnung der Abschlüsse in die Internationale Standardklassifikation des Bildungswesens (ISCED 2011) ist den zentralen Arbeitshilfen des FDZ (https://fdz.iab.de/anleitungen-und-hilfen/zentrale-arbeitshilfen) zu entnehmen. Je nach Quelle sind unterschiedliche Ausprägungen möglich.</p> <p>1) BeH</p> <p>Mit der Umstellung des Tätigkeitsschlüssels in 2011 ändern sich die möglichen Ausprägungen (siehe Kapitel 5.2.2). Davor gelten:</p> <ul style="list-style-type: none"> 5 Volks- / Hauptschule o./m. Abschluss, Mittlere Reife oder gleichwertige Schulbildung 8 Fachhochschulreife oder Abitur/Hochschulreife 9 Abitur / Hochschulreife <p>Danach gelten:</p> <ul style="list-style-type: none"> 1 Kein Hauptschulabschluss 4 Hauptschulabschluss / Volksschulabschluss 6 Mittlere Reife 8 Fachhochschulreife oder Abitur / Hochschulreife <p>Die Unterscheidungsmöglichkeiten im Sekundarbereich I wurden damit erweitert (1, 4, 6 statt vorher 5), die im Sekundarbereich II dagegen verringert (8 und 9 zusammengefasst unter 8).</p> <p>2) ASU, XASU, MTH, XMTH</p> <p>Für diese Quellen sind die folgenden Ausprägungen möglich:</p> <ul style="list-style-type: none"> 1 Kein Hauptschulabschluss 4 Hauptschulabschluss / Volksschulabschluss 6 Mittlere Reife 7 Fachhochschulreife

Kategorie	Beschreibung
	9 Abitur / Hochschulreife Sie sind zu Beginn der Arbeitsuche / Maßnahmeteilnahme gültig. In der Quelle XASU kann das Merkmal bei Ausbildungsplatzsuchenden auch den angestrebten Abschluss enthalten.
Qualitätshinweis	<p>1) BeH Der Füllgrad nimmt in der BeH über die Jahre kontinuierlich ab und liegt in jüngeren Jahren unter 2/3. Mit der Umstellung des Tätigkeitsschlüssels ist ein Bruch in den Daten verbunden, der über die reine Umstellung des Schlüssels hinausgeht, und evtl. durch Aktualisierungseffekte bei der Umstellung der Lohnbuchhaltungssoftware bei den meldenden Betrieben zu erklären ist.</p> <p>2) ASU, XASU, MTH, XMTH In den Quellen XASU und XMTH nimmt der Füllgrad kontinuierlich zu und liegt seit 2018 (XASU) bzw. seit 2015 (XMTH) über 75%. 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)

Kategorie	Beschreibung
Variablenlabel	Tagesentgelt / täglicher Leistungssatz
Variablenname	tentgelt
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH
Datentyp	Numerisch
Detailbeschreibung	<p>1) BeH In BeH-Sätzen weist das Merkmal jeweils das Bruttotagesentgelt eines Beschäftigten aus. Es wird berechnet aus dem vom Betrieb gemeldeten Zeitraumentgelt sowie der Dauer des Zeitraums der (ungesplitteten) Originalmeldung in Kalendertagen. Das Tagesentgelt ist in Euro angegeben.</p> <p>Bis 1998 meldeten Betriebe 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>

Kategorie	Beschreibung
	<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 in den zentralen Arbeitshilfen des FDZ (https://fdz.iab.de/anleitungen-und-hilfen/zentrale-arbeitshilfen).</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 Kapitel 6.5.15). 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>2) LeH</p> <p>Bei LeH-Datensätzen enthält das Merkmal den täglichen Leistungssatz, jeweils umgerechnet in Euro. Es ist zu berücksichtigen, dass der tägliche Leistungssatz bei Datensätzen mit einem Original-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. Weitere Informationen zu diesen Episoden finden sich in der Variablen „Abmeldegrund / Abgabegrund / Beendigungsgrund (grund)“.</p>

6.5.2 Tagesentgelt (inkl. Einmalzahlungen) (tentgelt_bonus)

Kategorie	Beschreibung
Variablenlabel	Tagesentgelt (inkl. Einmalzahlungen)
Variablenname	tentgelt_bonus
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch

Kategorie	Beschreibung
Detailbeschreibung	Das Merkmal Tagesentgelt (inkl. Einmalzahlungen) ist ein Zusatzmerkmal zum Merkmal Fehler! Verweisquelle konnte nicht gefunden werden. und enthält für BeH-Meldungen um Einmalzahlungen/Sonderzahlungen ergänzte und harmonisierte Angaben zum Tagesentgelt. Meldungen mit Abmeldegrund 54 (Meldung eines einmalig gezahlten Entgelts) müssen damit nicht mehr selbst integriert werden. Werte oberhalb von 97,5% des Modus der auf ganze Eurobeträge gerundeten Tagesentgelte vollzeitbeschäftigter Männer (pro Jahr und separat für Ost- und Westdeutschland) sind trunziert. Dieser Wert ist niedriger als die offizielle Beitragsbemessungsgrenze. Das Verfahren wird in Abschnitt 4.1. von Drechsler et al. (2023) beschrieben.
Qualitätshinweis	Das Merkmal ist nicht für alle Beobachtungen der Quelle BeH gefüllt. Zum einen werden Episoden mit Abmeldegrund 154 nicht berücksichtigt, zum anderen werden von Drechsler et al. (2023) Mehrfachbeschäftigungen zum gleichen Zeitraum im gleichen Betrieb vorab bereinigt und für die Generierung des Merkmals nicht verwendet. Im vorliegenden Datenprodukt sind diese Episoden noch enthalten.

6.5.3 Tagesentgelt (imputiert) (tentgelt_imp)

Kategorie	Beschreibung
Variablenlabel	Tagesentgelt (imputiert)
Variablenname	tentgelt_imp
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal Tagesentgelt (imputiert) ist ein Zusatzmerkmal zum Merkmal Tagesentgelt/täglicher Leistungssatz (tentgelt) und enthält für BeH-Meldungen von in Vollzeit Beschäftigten im Alter zwischen 14 und 64 Jahren (ohne Auszubildende) imputierte Tagesentgelte, wenn diese wegen Überschreitens der Beitragsbemessungsgrenze in den Meldedaten zensiert sind. Das Merkmal baut auf dem Merkmal Tagesentgelt (inkl. Einmalzahlungen) (tentgelt_bonus) auf. Tagesentgelte dieser Subpopulation, die die Beitragsbemessungsgrenze nicht überschreiten, werden aus „tentgelt_bonus“ übernommen. Details zum Imputationsverfahren sind in Drechsler et al. (2023) bzw. Drechsler/Ludsteck (2025) beschrieben.
Qualitätshinweis	Drechsler et al. (2023, Kapitel 5) warnen davor, das Merkmal ohne weitere Prüfschritte in Analysen zu verwenden. Insbesondere sollte immer geprüft werden, wie hoch der Anteil zensierter Entgeltangaben im jeweiligen Zielmodell ist, und ob alle Regressoren des Zielmodells entweder direkt oder durch ausreichend gute Proxies auch Eingang in das Imputationsmodell gefunden haben. Es wird daher empfohlen, sich vor einer Nutzung mit den Annahmen und Grenzen des Verfahrens vertraut zu machen.

6.5.4 Beruf – ausgeübte / letzte Tätigkeit (KldB 1988) (beruf)

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

Kategorie	Beschreibung
Detailbeschreibung	<p>1) BeH Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V). Der Arbeitgeber verschlüsselt die Tätigkeit 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. Beschäftigungsmeldungen mit einem Enddatum nach dem 30.11. 2011 werden nach dem neuen Tätigkeitsschlüssel 2010 (KldB2010) gemeldet (siehe Abschnitt 5.2.2). Diese Werte werden durch einen Schwerpunktumstieg auf die KldB1988 zurückgeschlüsselt. Dabei kommt es zu Unschärfen.</p> <p>2) ASU, XASU, MTH Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1).</p>
Qualitätshinweis	<p>1) BeH In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf. Mit der Umstellung des Tätigkeitsschlüssels ist ein Bruch in den Daten verbunden, der über die reine Umstellung des Schlüssels hinausgeht, und evtl. durch Aktualisierungseffekte bei der Umstellung der Lohnbuchhaltungssoftware bei den meldenden Betrieben zu erklären ist.</p> <p>2) ASU, MTH Das Merkmal ist nach 2011 nur noch dann gefüllt, wenn die letzte Tätigkeit noch nach dem alten Tätigkeitsschlüssel gemeldet wurde. Der Füllgrad fällt daher über die Jahre kontinuierlich ab. Nach 2016 liegt er unter 10%.</p> <p>3) XASU Das Merkmal Beruf ist nahezu für den gesamten zur Verfügung stehenden Zeitraum nicht gefüllt.</p>

6.5.5 Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010), 3-Steller (beruf2010_3)

Kategorie	Beschreibung
Variablenlabel	Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010), 3-Steller
Variablenname	beruf2010_3
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, ASU, XASU, MTH
Datentyp	Numerisch
Detailbeschreibung	<p>1) BeH 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>

Kategorie	Beschreibung
	<p>Der Arbeitgeber verschlüsselt die Tätigkeit der Beschäftigten oder des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2021). 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 – ausgeübte / letzte Tätigkeit (KldB 2010) (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 Kapitel 5.2.2). 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>Die KldB2010 wurde 2020 aktualisiert. Dabei wurden zwei neue Berufsuntergruppen (4-Steller) und 14 neue Berufsgattungen (5-Steller) geschaffen. Zudem wurden eine Berufsuntergruppe und eine Berufsgattung innerhalb der Systematik umgezogen. In der Quelle BeH sind die Änderungen ab 2020 enthalten.</p> <p>2) LeH, ASU, XASU, MTH</p> <p>Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1). Für diese Quellen sind die Aktualisierung der Klassifikation von 2020 auch rückwirkend eingepflegt, sodass sie schon ab 2006 zu finden sind.</p> <p>3) alle Quellen</p> <p>Die Berufsgruppe lässt sich über die Berufshauptgruppe (erste zwei Stellen des Schlüssels) den Berufssektoren und Berufssegmenten nach Klassifikation der Bundesagentur für Arbeit und des IAB zuordnen. Eine Einordnung ist den zentralen Arbeitshilfen des FDZ (https://fdz.iab.de/anleitungen-und-hilfen/zentrale-arbeitshilfen) zu entnehmen.</p>
Qualitätshinweis	<p>1) BeH</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. Mit der Umstellung des Tätigkeitsschlüssels ist ein Bruch in den Daten verbunden, der über die reine Umstellung des Schlüssels hinausgeht, und evtl. durch Aktualisierungseffekte bei der Umstellung der Lohnbuchhaltungssoftware bei den meldenden Betrieben zu erklären ist.</p> <p>2) LeH</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>3) XASU</p> <p>Das Merkmal ist für den gesamten zur Verfügung stehenden Zeitraum sehr schlecht gefüllt.</p>

6.5.6 Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010), 4-Steller (beruf2010_4)

Kategorie	Beschreibung
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

Kategorie	Beschreibung
Datentyp	Numerisch
Detailbeschreibung	<p>1) BeH</p> <p>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, 2021). 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 – ausgeübte / letzte Tätigkeit (KldB 2010) (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>Die KldB2010 wurde 2020 aktualisiert. Dabei wurden zwei neue Berufsuntergruppen (4-Steller) und 14 neue Berufsgattungen (5-Steller) geschaffen. Zudem wurden eine Berufsuntergruppe und eine Berufsgattung innerhalb der Systematik umgezogen. In der Quelle BeH sind die Änderungen ab 2020 enthalten.</p> <p>2) LeH, ASU, XASU, MTH</p> <p>Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1).</p> <p>Für diese Quellen sind die Aktualisierung der Klassifikation von 2020 auch rückwirkend eingepflegt, sodass sie schon ab 2006 zu finden sind.</p> <p>3) alle Quellen</p> <p>Die Berufsgruppe lässt sich über die Berufshauptgruppe (erste zwei Stellen des Schlüssels) den Berufssektoren und Berufssegmenten nach Klassifikation der Bundesagentur für Arbeit und des IAB zuordnen. Eine Einordnung ist den zentralen Arbeitshilfen des FDZ (https://fdz.iab.de/anleitungen-und-hilfen/zentrale-arbeitshilfen) zu entnehmen. Die als Meister tätigen Beschäftigten können, abgesehen von wenigen Ausnahmen in den Verkehrs-, Handels- und personenbezogenen Dienstleistungsberufen (beruf2010_3: 514, 621, 632, 633, 713, 714, 732, 813, 831, 832, 921, 944 und 946), durch eine Kombination des Werts 9 an der vierten Stelle (Berufsuntergruppe) mit dem Wert 3 im Anforderungsniveau (niveau) identifiziert werden.</p>
Qualitätshinweis	<p>1) BeH</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. Mit der Umstellung des Tätigkeitsschlüssels ist ein Bruch in den Daten verbunden, der über die reine Umstellung des Schlüssels hinausgeht, und evtl. durch Aktualisierungseffekte bei der Umstellung der Lohnbuchhaltungssoftware bei den meldenden Betrieben zu erklären ist.</p> <p>2) LeH</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>

Kategorie	Beschreibung
	3) XASU Das Merkmal ist für den gesamten zur Verfügung stehenden Zeitraum sehr schlecht gefüllt.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird als Berufsangabe gemäß KldB2010 lediglich die Berufsgruppe (beruf2010_3) bereitgestellt.

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

Kategorie	Beschreibung
Variablenlabel	Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010)
Variablenname	Niveau
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, ASU, XASU, MTH
Datentyp	Numerisch
Detailbeschreibung	<p>1) BeH</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, 2021). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Das, an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit, wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt.</p> <p>Beschäftigungsmeldungen mit einem Enddatum vor dem 30.11.2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet. Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten.</p> <p>Die KldB2010 wurde 2020 aktualisiert. Dabei wurden zwei neue Berufsuntergruppen (4-Steller) und 14 neue Berufsgattungen (5-Steller) geschaffen. Zudem wurden eine Berufsuntergruppe und eine Berufsgattung innerhalb der Systematik umgezogen. In der Quelle BeH sind die Änderungen ab 2020 enthalten.</p> <p>2) LeH, ASU, XASU, MTH</p> <p>Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1).</p> <p>Für diese Quellen sind die Aktualisierung der Klassifikation von 2020 auch rückwirkend eingepflegt, sodass sie schon ab 2006 zu finden sind.</p>
Qualitätshinweis	<p>1) BeH</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. Mit der Umstellung des Tätigkeitsschlüssels ist ein Bruch in den Daten verbunden, der über die reine Umstellung des Schlüssels hinausgeht, und evtl. durch Aktualisierungseffekte bei der Umstellung der Lohnbuchhaltungssoftware bei den meldenden Betrieben zu erklären ist.</p> <p>2) LeH</p>

Kategorie	Beschreibung
	<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, wodurch die Qualität gestiegen ist.</p> <p>3) XASU</p> <p>Das Merkmal ist für den gesamten zur Verfügung stehenden Zeitraum sehr schlecht gefüllt.</p>

6.5.8 Teilzeit (teilzeit)

Kategorie	Beschreibung
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	<p>In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Um dieses Problem abzumildern, wurde am IAB eine Imputation der Arbeitszeit für den fraglichen Zeitraum vorgenommen. Nähere Informationen zum Verfahren können Ludsteck/Thomsen (2016) entnommen werden. Eine ähnliche Häufung tritt 1991 aufgrund der Wiedervereinigung auf. Hier wird keine Imputation vorgenommen.</p> <p>Mit der Umstellung des Tätigkeitsschlüssels ist ein Bruch in den Daten verbunden, der über die reine Umstellung des Schlüssels hinausgeht, und evtl. durch Aktualisierungseffekte bei der Umstellung der Lohnbuchhaltungssoftware bei den meldenden Betrieben zu erklären ist. Eine weiterführende Diskussion des Problems und ein möglicher Korrekturansatz finden sich in Fitzenberger/Seidlitz (2020).</p>

6.5.9 Stellung im Beruf und Arbeitszeit (stib)

Kategorie	Beschreibung
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</p>

Kategorie	Beschreibung
	<p>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 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 5.2.2) fiel diese Unterscheidung jedoch weg. Die Variable „stib“ ist daher nur für Meldungen gefüllt, die noch vor Einführung des neuen Tätigkeitsschlüssels abgegeben wurden. Es besteht jedoch die Möglichkeit, Tätigkeiten, die eine Meisterausbildung erfordern, anhand der KldB 2010 zu identifiziert. Weitere Informationen finden Sie unter der Variablen „Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_4)“.</p>
Qualitätshinweis	In 1991 kommt es aufgrund der Wiedervereinigung zu einer deutlichen Häufung der fehlenden Angaben.

6.5.10 Erwerbsstatus (erwstat)

Kategorie	Beschreibung
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>1) BeH</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>

Kategorie	Beschreibung
	<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. Diese übrigen Meldungen wurden der Personengruppe "Sozialversicherungspflichtig Beschäftigte ohne besondere Merkmale" (101) zugeordnet.</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>2) LeH</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. Mit Einführung des Arbeitslosengeld II 2005 entfallen die Leistungen Arbeitslosenhilfe, Unterhaltsgeld und Beiträge zur privaten Pflegeversicherung.</p> <p>3) LHG</p> <p>Für LHG-Sätze zeigt das Merkmal Erwerbsstatus an, ob die Person minderjährig erwerbsfähig, 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>4) ASU, XASU</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>5) MTH, XMTH</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	1) BeH

Kategorie	Beschreibung
	<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. Diese übrigen Meldungen wurden der Personengruppe "Sozialversicherungspflichtig Beschäftigte ohne besondere Merkmale" (101) zugeordnet.</p> <p>2) LHG</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>3) ASU, XASU</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>

6.5.11 Maßnahmeart – Gruppe (mass)

Kategorie	Beschreibung
Variablenlabel	Maßnahmeart – Gruppe
Variablenname	mass
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal zeigt die Maßnahmeart-Gruppe an. Dies ist die zweithöchste Hierarchiestufe der Maßnahmeart-Systematiken der Bundesagentur für Arbeit.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen bereitgestellt. Ansonsten wird lediglich die Maßnahmeart-Kategorie bereitgestellt (siehe Merkmal „ Fehler! Verweisquelle konnte nicht gefunden werden. “).

6.5.12 Gleitzone (gleitz)

Kategorie	Beschreibung
Variablenlabel	Gleitzone

Kategorie	Beschreibung
Variablenname	gleitz
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Detailbeschreibung	<p>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 Gleitzzone bzw. ab Juli 2019 in einem monatlichen Bruttoarbeitsentgelt von 400,01 bis 800,00 € "Übergangsbereich" handelt (sog. Midi-Job), für die der Arbeitnehmer nur einen reduzierten Beitrag zum Gesamtsozialversicherungsbeitrag leisten muss. Da Arbeitnehmende mit einem Entgelt in der Gleitzzone freiwillig den „normalen“ Sozialversicherungsbeitrag zahlen können, sind nicht automatisch alle Beschäftigungen mit einem entsprechen Entgelt als Gleitzzone gekennzeichnet.</p> <p>Die Zuordnung erfolgt über das Bruttoarbeitsentgelt und ändern sich im Zeitverlauf: 01.04.2003 bis 31.12.2012: 400,01 bis 800,00 Euro 01.01.2013 bis 30.06.2019: 450,01 bis 850,00 Euro 01.07.2019 bis 31.12.2022: 450,01 bis 1.300,00 Euro 01.10.2022 bis 31.12.2022: 520,01 bis 1.600 Euro Ab 01.01.2023: 520,01 bis 2.000 Euro</p> <p>Abweichende untere Grenzen existieren für Auszubildende und Beschäftigte in Freiwilligendiensten. Bei mehreren Beschäftigungsverhältnissen ist das insgesamt erzielte Arbeitsentgelt maßgebend.</p> <p>Die entsprechende gesetzliche Regelung besteht seit dem 01.04.2003.</p>

6.5.13 Leiharbeit (leih)

Kategorie	Beschreibung
Variablenlabel	Leiharbeit
Variablenname	leih
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Detailbeschreibung	<p>Das Merkmal gibt an, ob es sich bei der Beschäftigung der Person um ein Leiharbeitsverhältnis handelt. Die Meldung erfolgt durch den verleihenden Betrieb, somit werden die Beschäftigten dem verleihenden Betrieb zugeordnet. Das Merkmal entstammt dem Tätigkeitsschlüssel 2010 und ist nur für Meldungen mit einem Enddatum nach dem 30.11.2011 gefüllt.</p>
Qualitätshinweis	<p>In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Im Jahr 2012 beträgt die Missingquote nur noch rund 3 % und sinkt bis 2019 auf unter 1%.</p>

6.5.14 Befristung (befrist)

Kategorie	Beschreibung
Variablenlabel	Befristung
Variablenname	befrist
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

Kategorie	Beschreibung
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 % und sinkt bis 2019 auf unter 1%.

6.5.15 Abmeldegrund / Abgabegrund / Beendigungsgrund (grund)

Kategorie	Beschreibung
Variablenlabel	Abmeldegrund / Abgabegrund / Beendigungsgrund
Variablenname	grund
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, LHG, ASU, XASU
Datentyp	Numerisch
Detailbeschreibung	<p>1) BeH</p> <p>In BeH-Meldungen kennzeichnet der Abgabegrund den Anlass, aus dem der Arbeitgeber den Sozialversicherungsträgern die betreffende Beschäftigungsmeldung übermittelt hat. Allerdings werden in die IEB nicht alle im Rahmen des Meldeverfahrens 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) mit einer führenden „1“ kodiert. Um die Originalausprägungen zu erhalten, muss der Wert „100“ von den Ausprägungen abgezogen werden.</p> <p>2) LeH</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>Ab 2007 werden außerdem die Beendigungsgründe "Ende der Sperrzeit" (1183) und "Ende der Unterbrechungszeit" (1184) gesetzt, wenn es sich bei einer Episode um einen Sperr- oder Unterbrechungszeit handelt. Dies gilt jedoch nicht, wenn diese Episode den gesamten Anspruchszeitraum abschließt; in diesem Fall wird der Original-Beendigungsgrund des Anspruchs beibehalten. Sperr- und Unterbrechungszeiträume können jedoch über einen täglichen Leistungssatz von 0 Euro (siehe Variable Tagesentgelt / täglicher Leistungssatz (tentgelt)) identifiziert werden. Zur Unterscheidung von Sperr- und Unterbrechungszeiträume kann zusätzlich das Merkmal „restanspruch“ herangezogen werden: In der Regel verringert sich die Anspruchszeiten bei Unterbrechungszeiträume nicht, bei Sperrzeiten jedoch schon (siehe Variable Restanspruch/geplante Dauer (restanspruch)).</p>

Kategorie	Beschreibung
	<p>Die Ausprägungen dieses Merkmals sind nur für den fachlich bzw. gesetzlich vorgesehenen Gültigkeitszeitraum gefüllt. Kommen Ausprägungen außerhalb ihres Gültigkeitszeitraums vor, werden sie auf Ersatzwerte umgeschlüsselt {Wertebereich 1173-1181}, weil ihre Bedeutung dann nicht sicher geklärt werden kann.</p> <p>3) LHG</p> <p>Bei LHG-Datensätzen ist der SGB-II-Beendigungsgrund 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. Ändert sich die Zusammensetzung einer Bedarfsgemeinschaft, werden die Sätze für alle Mitglieder in einer Bedarfsgemeinschaft zu diesem Datum abgeschlossen und neu begonnen. Der SGB-II-Beendigungsgrund ist jedoch nur für die Personen gefüllt, für die ein Beendigungsgrund vorliegt. Bei anderen Bedarfsgemeinschaftsmitgliedern enthält der SGB-II-Beendigungsgrund missings (.z). Das Merkmal ist genau zum Ende des Originalsatzes gültig.</p> <p>4) ASU</p> <p>Bei ASU-Sätzen ist der Abmeldegrund 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 Regeln umgeschlüsselt werden: 229 auf 260, 230 auf 260, 231 auf 261, 232 auf 260, 233 auf 260, 234 auf 260, 235 auf 260, 236 auf 261, 237 auf 266, 238 auf 266, 239 auf 271, 240 auf 269, 242 auf 265, 243 auf 270, 244 auf 274, 245 auf 277, 246 auf 267, 247 auf 267, 248 auf 278, 249 auf 269, 250 auf 275, 251 auf 274, 252 auf 276, 253 auf 268, 254 auf 278</p> <p>5) XASU</p> <p>Bei XASU-Sätzen ist der Abmeldegrund enthalten. In der XASU werden überschneidungsfreie Zeiträume bezogen auf einen zKT-Kunden abgebildet. Bei Änderungen bestimmter Merkmale wird für die XASU jeweils ein neuer Datensatz gebildet, davon sind folgende Bestandteil der SIAB:</p> <ul style="list-style-type: none"> • Wechsel des Arbeitssuche-Status • Wechsel der Verfügbarkeit • Wechsel des SGB-II-Trägers • Wechsel des Wohnortes <p>Wird die vom Grundsicherungsträger gemeldete Episode aus anderen Gründen künstlich gesplittet (siehe Beginndatum der Episode (begepi)), wird als Abmeldegrund "DS durch Splittung erzeugt" eingetragen.</p>
Qualitätshinweis	<p>1) BeH</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>2) LHG, XASU</p> <p>Der Anteil valider Werte (Füllgrad) des Abgabegrundes ist in der Quelle LHG über alle Jahre sehr niedrig (< 20%).</p>

6.5.16 Erwerbsstatus vor Arbeitsuche (estatvor)

Kategorie	Beschreibung
Variablenlabel	Erwerbsstatus vor Arbeitsuche
Variablenname	estatvor
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, XASU
Datentyp	Numerisch
Detailbeschreibung	Die Variable zeigt für ASU- und XASU-Sätze den Erwerbsstatus vor der Arbeitsuche an. Wird ein Datensatz innerhalb der ASU/XASU künstlich gesplittet (etwa bei einem Rechtskreiswechsel zwischen SGB III und SGB II), wird der Zugangsgrund „DS durch Splitting erzeugt“ vergeben. Dies gilt jedoch nicht für das Episodensplitting innerhalb der SIAB. Die Angabe bezieht sich daher nicht notwendigerweise auf das Beginndatum der Episode, sondern auf den Beginn des Originalzeitraums.
Qualitätshinweis	1) ASU Mit dem Umstieg von coArb zu VerBIS in 2005/2006 (siehe Kapitel 5.6) ist ein deutlicher Bruch in den Merkmalsausprägungen zu verzeichnen, da ein einheitliches Mapping nicht möglich ist.

6.5.17 Status nach Arbeitsuche / SGB-II-Ausschlussgrund / Verfügbarkeit (estatnach)

Kategorie	Beschreibung
Variablenlabel	Status nach Arbeitsuche / SGB-II-Ausschlussgrund / Verfügbarkeit
Variablenname	estatnach
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU
Datentyp	Numerisch
Detailbeschreibung	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: <ul style="list-style-type: none"> • 1000er: Maßnahme (geförderte Beschäftigung) • 2000er: ungeförderte Beschäftigung • 3000er: in Ausbildung etc. • 4000er: in Selbständigkeit • 5000er: in Sperre • 6000er: in Sonstige
Qualitätshinweis	Mit dem Umstieg von coArb zu VerBIS in 2005/2006 (siehe Abschnitt 5.6) ist ein deutlicher Bruch in den Merkmalsausprägungen zu verzeichnen, da ein einheitliches Mapping nicht möglich ist. Auch sonst zeigen sich je nach Merkmalsausprägung im Zeitverlauf Auffälligkeiten in den relativen Häufigkeiten. Bei Auswertungen ist daher besondere Vorsicht geboten.

6.5.18 Integrationsprognose (ipo)

Kategorie	Beschreibung
Variablenlabel	Integrationsprognose
Variablenname	ipo
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

Kategorie	Beschreibung
Herkunft	ASU, MTH
Datentyp	numerisch
Detailbeschreibung	<p>Das Merkmal enthält die im Erstkontakt von der Vermittlungs-, Beratungs- und Integrationsfachkraft (bzw. dem Lotsen im SGB III) eingeschätzte Integrationsprognose. Die Integrationsprognose ist das Ergebnis der Einschätzung zur Marktnähe der Kundinnen und Kunden. "Marktnah" gilt dann, wenn eine Integration innerhalb von sechs Monaten zu erwarten ist, "nicht marktnah", wenn die Integration voraussichtlich erst nach mehr als sechs Monaten gelingt. Erkennbare Handlungsbedarfe, die innerhalb von sechs Monaten behoben werden können, schließen Marktnähe nicht aus.</p> <p>Die Integrationsprognose "integriert, aber hilfebedürftig" ist nur gültig für Bewerberinnen und Bewerber aus dem Rechtskreis SGB II.</p> <p>Besteht ausschließlich ein Vermittlungswunsch in Ausbildung, ist keine Integrationsprognose festzulegen.</p> <p>Dieses Merkmal steht ab Ende 2016 zur Verfügung und ist Nachfolger des Merkmals Profillage (profil). Für die Zeit davor werden die Angaben aus der Profillage in die Integrationsprognose wie folgt umgeschlüsselt:</p> <p>Die alten Ausprägungen 2 (Marktprofil), 3 (Aktivierungsprofil) und 4 (Förderprofil) werden zur neuen Ausprägung 1 (Marktnah), die alten Ausprägungen 5 (Entwicklungsprofil), 6 (Stabilisierungsprofil) und 7 (Unterstützungsprofil) zur neuen Ausprägung 2 (nicht marktnah), die alte Ausprägung 8 (Zuordnung nicht erforderlich) zu 3 (Zuordnung nicht erforderlich) und 9 (Integriert, aber hilfebedürftig) zu 4 (Integriert, aber hilfebedürftig) sowie 1 (Noch nicht festgelegt) zu .z (Keine Angabe).</p>
Qualitätshinweis	Das Merkmal wurde 2016 eingeführt, aber bis 2006 aus anderen Merkmalen umgeschlüsselt. Die Qualität wird von der BA-Statistik seit 2010 als verlässlich eingestuft.

6.5.19 Art der Kündigung der letzten Tätigkeit (art_kuend)

Kategorie	Beschreibung
Variablenlabel	Art der Kündigung der letzten Tätigkeit
Variablenname	art_kuend
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, MTH
Datentyp	Numerisch
Detailbeschreibung	<p>Diese Variable ist für Beschäftigungsmeldungen nicht gefüllt.</p> <p>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.</p>
Qualitätshinweis	Der Anteil der fehlenden Werte ist in dieser Variable besonders hoch und liegt durchschnittlich bei etwa 50%.

6.5.20 Arbeitszeit des Stellengesuchs (arbeitszeit)

Kategorie	Beschreibung
Variablenlabel	Arbeitszeit des Stellengesuchs
Variablenname	arbeitszeit
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

Kategorie	Beschreibung
Herkunft	ASU, MTH
Datentyp	Numerisch
Detailbeschreibung	Arbeitsuchende geben beim Vermittlungsvorgang an, welchem zeitlichen Umfang die gesuchte Stelle entsprechen sollte.

6.5.21 Restanspruch/geplante Dauer (restanspruch)

Kategorie	Beschreibung
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>1) LeH Das Merkmal enthält den Anspruch, der nach Ende des aktuellen Leistungsbezugs übrigbleibt. Endet der Leistungsbezug vor dem Ende der maximalen Anspruchsdauer (z. B. durch Arbeitsaufnahme), bleibt ein Restanspruch mit der Dauer des nicht ausgeschöpften Anspruchs bestehen. Entsteht innerhalb von fünf Jahren ein neuer Anspruch, wird die Dauer des Restanspruchs zur neuen Anspruchsdauer addiert. Allerdings ist hierbei die Höchstdauer für das jeweilige Lebensalter die Obergrenze. Entsteht kein neuer Anspruch, so kann der Restanspruch auf Antrag innerhalb von vier Jahren für Leistungen verwendet werden.</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>2) MTH Das Merkmal enthält die geplante Dauer der Maßnahme.</p>
Qualitätshinweis	<p>1) LeH 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.</p> <p>Darüber hinaus kann die Angabe zur Anspruchsdauer in Einzelfällen die aus den gesetzlichen Vorgaben abzuleitende individuelle Obergrenze übersteigen.</p> <p>2) MTH In sehr wenigen Fällen enthält die geplante Dauer der Maßnahme negative Werte oder Werte größer 5 Jahre. Diese Fälle werden auf missing (.z) umgesetzt.</p>

6.5.22 Trägerart (traeger)

Kategorie	Beschreibung
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 mit den Grenzen der Kreise übereinstimmen müssen. Gleiches gilt für die Arbeitsagenturen.</p>

6.5.23 Beginndatum der Arbeitslosigkeit (alo_beg)

Kategorie	Beschreibung
Variablenlabel	Beginndatum der Arbeitslosigkeit
Variablenname	alo_beg
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, XASU
Datentyp	Numerisch
Detailbeschreibung	<p>Das Merkmal ist für alle Quellen gefüllt und berücksichtigt ALO-Spell aus ASU- und XASU für die Berechnung. 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"> • jede Lücke kleiner oder gleich 7 Tage • Krankheitslücken bis zu 42 Tagen

6.5.24 Dauer der Arbeitslosigkeit (alo_dau)

Kategorie	Beschreibung
Variablenlabel	Dauer der Arbeitslosigkeit
Variablenname	alo_dau
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, XASU
Datentyp	Numerisch
Detailbeschreibung	<p>Das Merkmal ist für alle Quellen gefüllt und berücksichtigt ALO-Spell aus ASU- und XASU für die Berechnung. 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"> • jede Lücke kleiner oder gleich 7 Tage

Kategorie	Beschreibung
	<ul style="list-style-type: none"> • Krankheitslücken bis zu 42 Tagen <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)

Kategorie	Beschreibung
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>Für die Quellen LHG und XASU gilt der Wohnort für den gesamten Zeitraum der Originalbeobachtung. Bei BeH, 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 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 31.12.2023 umgeschlüsselt, d. h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2023 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 in der BeH 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>

Kategorie	Beschreibung
	Im Jahr 2015 kommt es in den BeH-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.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird als Regionalangabe lediglich das Bundesland (wo_bula) bereitgestellt.

6.6.2 Wohnort – Bundesland (wo_bula)

Kategorie	Beschreibung
Variablenlabel	Wohnort - Bundesland
Variablenname	wo_bula
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Hierarchie	Bundesland Kreis
Detailbeschreibung	Dieses Merkmal ist eine Aggregation aus dem Merkmal Wohnort - Kreis auf die 16 deutschen Bundesländer. 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).

6.6.3 Wohnort – Arbeitsagentur (wo_aa)

Kategorie	Beschreibung
Variablenlabel	Wohnort - Arbeitsagentur
Variablenname	wo_aa
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Hierarchie	Regionaldirektion Arbeitsagentur
Detailbeschreibung	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 BeH, LeH, ASU, MTH und XMTH enthält das Merkmal den Wohnort zum 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.

Kategorie	Beschreibung
	<p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Angaben für alle Quellen auf den Gebietsstand vom 31.12.2023 umgeschlüsselt, d. h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Agenturbezirk nach den Grenzen, welche die Bezirke zum 31.12.2023 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)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
Variablenlabel	WZ73 3-Steller
Variablenname	w73_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabteilung (1-Steller) Wirtschaftsgruppe (2-Steller) Wirtschaftsklasse (3-Steller)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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 widergespiegelt werden.</p>
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w93_3) bereitgestellt.

6.7.3 WZ93 3-Steller (w93_3)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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>

6.7.4 WZ03 5-Steller (w03_5)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
Variablenlabel	w93_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w93_3_gen

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
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>

Kategorie	Beschreibung
	Weitere Informationen zur Klassifikation WZ08 finden sich bei der Variablenbeschreibung von w08_3.

6.7.13 Art der Vervollständigung w08_3 (group_w08_3)

Kategorie	Beschreibung
Variablenlabel	Art der Vervollständigung w08_3
Variablenname	group_w08_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	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. Eine ausführliche Beschreibung zur Vorgehensweise findet sich bei Eberle et al., 2011.

6.7.14 Jahr erstes Auftreten Betriebsnummer (grd_jahr)

Kategorie	Beschreibung
Variablenlabel	Jahr erstes Auftreten Betriebsnummer
Variablenname	grd_jahr
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Bei dieser Variable 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 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. 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.

6.7.15 Jahr letztes Auftreten Betriebsnummer (lzt_jahr)

Kategorie	Beschreibung
Variablenlabel	Jahr letztes Auftreten Betriebsnummer
Variablenname	lzt_jahr
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch

Kategorie	Beschreibung
Detailbeschreibung	<p>Bei dieser Variable handelt es sich um das letzte Auftreten der Nummer des Betriebes im Datensatz (vgl. Bender et al., 1996). Endet die Existenz einer Betriebsnummer im BHP bereits vor dem Jahr 2023, 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).</p> <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.</p>

6.7.16 Anzahl Beschäftigte gesamt (az_ges)

Kategorie	Beschreibung
Variablenlabel	Anzahl Beschäftigte gesamt
Variablenname	az_ges
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
Detailbeschreibung	<p>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.</p>

6.7.17 Anzahl Vollzeit (Normalbeschäftigte + sonstige) (az_vz)

Kategorie	Beschreibung
Variablenlabel	Anzahl Vollzeit (Normalbeschäftigte + sonstige)
Variablenname	az_vz
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
Detailbeschreibung	<p>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.</p>

6.7.18 Anzahl geringfügig Beschäftigte (az_gf)

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

Kategorie	Beschreibung
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.

6.7.19 Mittelwert imputiertes Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)

Kategorie	Beschreibung
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. 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 Ganzer et al. 2024). 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>

6.7.20 Arbeitsort - Kreis (ao_kreis)

Kategorie	Beschreibung
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 unités territoriales statistiques) der Europäischen Union entsprechen die Kreise der Hierarchieebene NUTS 3.

Kategorie	Beschreibung
	Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben auf den Gebietsstand vom 31.12.2023 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Betriebsortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2023 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.
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.

6.7.21 Arbeitsort - Bundesland (ao_bula)

Kategorie	Beschreibung
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)

Kategorie	Beschreibung
Variablenlabel	Vorliegen administrativer Personendaten
Variablenname	match_admin
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	Record Linkage
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal ist nur in der Verknüpfungsdatei enthalten. Es zeigt für jede befragte Person in SOEP-CMI an, ob zu dieser Person 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 Verknüpfungsart (match_typ)

Kategorie	Beschreibung
Variablenlabel	Verknüpfungsart
Variablenname	match_typ
Kategorie	Verknüpfungsbezogenes Merkmal

Kategorie	Beschreibung
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 Teilpopulation (group)

Kategorie	Beschreibung
Variablenlabel	Teilpopulation
Variablenname	group
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	SOEP-CMI
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal ist nur in der Verknüpfungsdatei enthalten. Damit ist eine einfachere Abgrenzung unterschiedlicher Teilpopulationen bzw. Gruppen von Stichproben in SOEP-CMI möglich.

7 Literatur

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. URL: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Methodik-Qualitaet/Methodenberichte/Beschaeftigungsstatistik/Generische-Publikationen/MB-Arbeitszeit-ausgeuebte-Taetigkeit-Schul-und-Berufsabschluss.pdf?__blob=publicationFile&v=4 (Zugriff: 12.03.2026).

Borgman, C.L., & Siegfried, S.L. (1992). Getty's Synonym™ and its cousins: A survey of applications of personal name-matching algorithms. Journal of the American Society for Information Science, Jg. 43 (7), S. 459-476. DOI: 10.1002/(SICI)1097-4571(199208)43:7<459::AID-ASI1>3.0.CO;2-D

Brücker, H., Kosyakova, Y., Rother, R., Zinn, S., Eckhard, Maddox, Niehues, Schwanhäuser, S., Siegert, M., Tanis, Trübswetter, P., Goebel, J., Grabka, M., Schröder, C., Bartels, C., Beckmannshagen, M., Franken, A., Gerike, M., Griese, F., Halbmeier, C., Kara, S., Krause, P., Liebau, E., Nebelin, J., Petrenz, M., Satilmis, S., Siegers, R., Steinhauer, H. W., Wenzig, W., Zimmermann, S. (2023) IAB-BAMF-SOEP-Befragung Geflüchteter – Version 1622 v1. DOI: 10.5684/soep.iab-bamf-soep-mig.2022

- Brücker, H., Liebig, S., Goebel, J., Grabka, Markus M., Schröder, C., Zinn, S., Bartels, C., Fedorets, A., Franken, A., Gerike, M., GRIESE, F., Jacobsen, J., Kara, S., Krause, P., Kröger, H., Liebau, E., Metzinger, M., Nebelin, J., Pagel, L., Petrenz, M., Richter, D., Schacht, D., Schmelzer, P., Schmitt, C., Schnitzlein, D., Schupp, J., Siegers, R., Steinhauer, H. W., Trübswetter, P., Wenzig, K., Zimmermann, S. (2024): "IAB-SOEP Migrationsstichprobe Befragungsdaten – Version 1322 v1". DOI: 10.5684/soep.iab-soep-mig.2022
- 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. URL: <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Wirtschaftszweige/Klassifikation-der-Wirtschaftszweige-1973-2003/Generische-Publikationen/WS-1973-Excel.xlsx? blob=publicationFile&v=6> (Zugriff: 12.03.2026).
- Bundesagentur für Arbeit (2021): Klassifikation der Berufe 2010 – überarbeitete Fassung 2020 Band 1: Systematischer und alphabetischer Teil mit Erläuterungen. Nürnberg. URL: <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Printausgabe-KldB-2010-Fassung2020/Generische-Publikationen/KldB2010-PDF-Version-Band1-Fassung2020.pdf? blob=publicationFile&v=21> (Zugriff: 12.03.2026).
- Bundesanstalt für Arbeit (1988). Klassifikation der Berufe–Systematisches und Alphabetisches Verzeichnis der Berufsbenennung, Nürnberg. URL: <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB1975-1992/Generische-Publikationen/KldB1988-Systematisches-und-alphabetisches-Verzeichnis-der-Berufsbenennungen-PDF.pdf? blob=publicationFile&v=12> (Zugriff: 12.03.2026).
- Christen, P. (2012). Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection, Berlin, Springer.
- Cornesse, C., Gerike, M., Goebel, J., McKone Leonard, M., Liebig, S., Richter, D., Schröder, C., Wenzig, K., & Zinn, S. (2025). SOEP-Innovationssample (SOEP-IS), Daten der Jahre 1998-2023. DOI: 10.5684/soep.is.2023
- Cramer, U. (1985). Probleme der Genauigkeit der Beschäftigtenstatistik. Allgemeines Statistisches Archiv 69, S. 56-68.
- Drechsler, J. & Ludsteck, J. (2025): Imputation strategies for rightcensored wages in longitudinal datasets. Journal for Labour Market Research Jg. 59 (1). DOI: 10.1186/s12651-025-00410-4
- Drechsler, J., Ludsteck, J. & Moczall, A. (2023): Imputation der rechtszensierten Tagesentgelte für die BeH. FDZ-Methodenreport 05/2023 (de), Nürnberg, DOI:10.5164/IAB.FDZM.2305.de.v1
- 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.

- Fitzenberger, B. & Seidlitz, A. (2020). The 2011 break in the part-time indicator and the evolution of wage inequality in Germany. *Journal for Labour Market Research*, Jg. 54 (1). DOI: 10.1186/s12651-019-0265-0.
- Ganzer, A., Schmucker, A. & Wolter, S. (2024): Betriebs-Historik-Panel 1975–2023. FDZ-Datenreport 09/2024 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.2409.de.v1
- Goebel, J., Grabka, M. M., Schröder, C., Zinn, S., Bartels, C., Beckmannshagen, M., Franken, A., Gerike, M., Griese, F., Halbmeier, C., Kara, S., Krause, P., Liebau, E., Nebelin, J., Petrenz, M., Satilmis, S., Siegers, R., Steinhauer, H. W., Süttmann, F., Wenzig, K., Dräger, J., Gauer, M., Tchokni, Y., Saalbach, C., Labarca Pinto, D. & Deutsches Institut für Wirtschaftsforschung (DIW Berlin). (2025). Sozio-oekonomisches Panel (SOEP), Daten der Jahre 1984-2023 (SOEP-Core v40.1, EU-Edition - Update). DOI: 10.5684/soep.core.v40.1eu
- Institut für Arbeitsmarkt- und Berufsforschung (2024): Integrierte Erwerbsbiografien (IEB) 17.01.00-202312, Nürnberg.
- Lochner, B. & Wolter, Stefanie (2025): AKM effects for German labour market data 1985-2023. FDZ-Methodenreport 03/2025 (en), Nürnberg. DOI: 10.5164/IAB.FDZM.2503.en.v1
- 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.
- Richter, D., & Schupp, J. (2015). The SOEP Innovation Sample (SOEP IS). *Schmollers Jahrbuch*, Jg. 135 (3), S. 389-399. DOI: 10.3790/schm.135.3.389
- 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., & vom Berge, P. (2025): Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975 - 2023 v2. FDZ Datenreport, 02/2025 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.2502.de.v2
- 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: 12.03.2026).
- Statistisches Bundesamt (2003). Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-2003-erlaeuterung.pdf?_blob=publicationFile (Zugriff: 12.03.2026).
- 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: 12.03.2026).
- Statistisches Bundesamt (2024): Staats- und Gebietssystematik, Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluessel_pdf.pdf?_blob=publicationFile&v=5 (Zugriff: 12.03.2026).

- Statistik der Bundesagentur für Arbeit (2021): Grundlagen: Methodenbericht – Einführung der „Klassifikation der Berufe 2010 – überarbeitete Fassung 2020“ in die Arbeitsmarktstatistiken, Nürnberg. URL: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Methodik-Qualitaet/Methodenberichte/Uebergreifend/Generische-Publikationen/Methodenbericht-Klassifikation-Berufe-ueberarbeitete-Fassung.pdf?_blob=publicationFile&v=4 (Zugriff: 12.03.2026).
- Stüber, H., Dauth, W. & Eppelsheimer, J. (2023): A guide to preparing the sample of integrated labour market biographies (SIAB, version 7519 v1) for scientific analysis. Journal for Labour Market Research, Volume 57, Article number: 7. DOI: 10.1186/s12651-023-00335-w
- 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.
- Weitz, J., Ruland, M., Weiß, T., Hess, D., Liebich, L., Müller, T., Eisleb, J., Ponert, M. (2025). SOEP-Core – 2023: Report of Survey Methodology and Fieldwork. SOEP Survey Papers 1495: Series B. Berlin: DIW Berlin.
- 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 Labels, Auszählungen und fehlende Werte

Der Online-Anhang zum Datenreport stellt detaillierte Übersichten zu den einzelnen Merkmalen im Datenprodukt bereit. Er enthält Labellisten der einzelnen Ausprägungen, Auszählungen sowie Übersichten über den Anteil fehlender Werte (missings) getrennt nach Datenquellen. Sie finden den Anhang in gesonderten Dateien in deutscher und englischer Sprache unter:

- https://doku.iab.de/fdz/reporte/2026/DR_02-26_v1_frequencies_labels_de.zip
- https://doku.iab.de/fdz/reporte/2026/DR_02-26_v1_frequencies_labels_en.zip

8.2 Abkürzungsverzeichnis

ALG II	Arbeitslosengeld II
ARGE	Arbeitsgemeinschaft
ASU	Arbeitsuchendenhistorik
A2LL	Arbeitslosengeld II – Leistungen zum Lebensunterhalt
BA	Bundesagentur für Arbeit
BeH	Beschäftigtenhistorik
BfA	Bundesversicherungsanstalt für Angestellte
BG	Bedarfsgemeinschaft
BHP	Betriebs-Historik-Panel
BMAS	Bundesministerium für Arbeit und Soziales
coArb	Computerunterstützte Arbeitsvermittlung (operatives Verfahren zur Verwaltung der Vermittlung (Altverfahren))
DEÜV	Verordnung über die Erfassung und Übermittlung von Daten für die Träger der Sozialversicherung – Datenerfassungs- und -übermittlungsverordnung –
DEVO	Zweite VO über die Erfassung von Daten für die Träger der Sozialversicherung und für die BA – Datenerfassungs-Verordnung –
DIW Berlin	Deutsches Institut für Wirtschaftsforschung e.V.
DÜVO	Zweite VO über die Datenübermittlung auf maschinell verwertbaren Datenträgern im Bereich der Sozialversicherung und der BA – Datenübermittlungs-Verordnung–
EDV	Elektronische Datenverarbeitung
FDZ	Forschungsdatenzentrum
FELEG	Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit
gAw	Träger mit getrennter Aufgabenwahrnehmung
gE	Gemeinsame Einrichtungen
gT	getrennte Trägerschaft

IAB	Institut für Arbeitsmarkt- und Berufsforschung
IAB-FDZ	Forschungsdatenzentrum der Bundesagentur für Arbeit am IAB
IAB-SOEP-MIG	IAB-SOEP Migrationsstichprobe
IAB-BAMF-SOEP	IAB-BAMF-SOEP Befragung von Geflüchteten
IEB	Integrierte Erwerbsbiografien
ISIC	International Standard Industrial Classification of All Economic Activities
IZA	Forschungsinstitut zur Zukunft der Arbeit
KldB	Klassifikation der Berufe
LeH	Leistungsempfängerhistorik
LHG	Leistungshistorik Grundsicherung
LVA	Landesversicherungsanstalt
MTH	Maßnahmeteilnahmehistorik
NACE	Nomenclature générale des activités économiques dans les communautés européennes
NUTS	Nomenclature des unités territoriales statistique
SGB	Sozialgesetzbuch
SIAB	Stichprobe der Integrierten Arbeitsmarktbiografien
SOEP	Sozio-oekonomische Panel
SOEP-CMI	SOEP-Core-, IAB-SOEP Migrationsstichprobe, IAB-BAMF-SOEP Befragung von Geflüchteten und SOEP-Innovationssample
VerBIS	Vermittlungs- und Beratungsinformationssystem
XASU	Arbeitsuchendenhistorik aus XSozial-BA-SGB II
XMTH	Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II
zkT	Zugelassene kommunale Träger

Impressum

FDZ-Datenreport 02|2026

Veröffentlichungsdatum

18. März 2026

Herausgeber

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

Nutzungsrechte

Diese Publikation ist unter folgender Creative-Commons-Lizenz veröffentlicht:
Namensnennung - Weitergabe unter gleichen Bedingungen 4.0 International (CC BY-SA 4.0)
<https://creativecommons.org/licenses/by-sa/4.0/deed.de>

Bezugsmöglichkeit dieses Dokuments

https://doku.iab.de/fdz/reporte/2026/DR_02-26.pdf

Dokumentation Version

SOEP-CMI-ADIAB 7523_DE_v1_dok1, DOI: 10.5164/IAB.FDZD.2602.de.v1

Datensatz Version

SOEP-CMI-ADIAB 7523 v1, DOI: 10.5164/IAB.SOEP-CMI-ADIAB7523.de.en.v1

Bezugsmöglichkeit aller Veröffentlichungen der Reihe „FDZ-Datenreport“

<https://fdz.iab.de/forschung/publikationen/fdz-datenreporte>

Website

<https://fdz.iab.de>

Rückfragen zum Inhalt

Manfred Antoni

Telefon: 0911 179-5645

E-Mail: manfred.antoni@iab.de