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

FDZ-DATENREPORT

Dokumentation zu Arbeitsmarktdaten

13|2022 DE Datenspeicher Gesellschaftliches Arbeitsvermögen verknüpft mit administrativen Daten des IAB (GAV-ADIAB) 1975-2019

Florian Zimmermann, Matthias Umkehrer, Kevin Ruf, Dana Müller, Stefan Seth

Bundesagentur für Arbeit

Datenspeicher Gesellschaftliches Arbeitsvermögen verknüpft mit administrativen Daten des IAB (GAV-ADIAB) 1975-2019

Florian Zimmermann (IAB), Matthias Umkehrer (IAB), Kevin Ruf (IAB), Dana Müller (IAB), Stefan Seth (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 ... 8

1.2.1 Datenzugang ... 8

1.2.2 Datenhaltung ... 9

1.2.3 Dateinamen der administrativen Daten und der GAV-Daten ... 13

1.2.4 Datenverknüpfen ... 14

1.3 Steckbrief ... 15

1.4 Merkmalsliste ... 17

1.5 Mengengerüst ... 21

2 Datenquellen ... 22

2.1 GAV-Daten ... 22

2.2 Administrative Daten aus den IEB ... 23

2.2.1 Beschäftigtenhistorik (BeH) ... 23

2.2.2 Leistungsempfängerhistorik (LeH) ... 24

2.2.3 Leistungshistorik Grundsicherung (LHG) ... 24

2.2.4 Arbeitsuchendenhistoriken (ASU/XASU) ... 25

2.2.5 Maßnahmeteilnahmehistoriken (MTH/XMTH) ... 25

3 Datenaufbereitung und Stichprobenziehung ... 26

3.1 GAV-Daten ... 26

3.1.1 Bereinigung und Aufbereitung ... 26

3.1.2 Verknüpfung der Personendaten ... 26

3.2 Administrative Daten aus den IEB ... 27

3.2.1 Bereinigungen und Aufbereitungen ... 27

3.2.2 SGB-II-Anonymisierung ... 30

3.2.3 Episodensplitting ... 30

3.3 Stichprobendesign und Analysepotenzial ... 31

3.4 Gewichtung ... 32

3.5 Fehlende Werte ... 33

4 Datenqualität und Probleme ... 33

4.1 GAV ... 33

4.2 Gesamte IEB ... 34

4.2.1 Lücken in den Erwerbsbiografien ... 34

4.2.2 Einführung des SGB II und spätere Trägerwechsel ... 36

4.3 Beschäftigtenhistorik ... 36

4.4 Leistungsempfängerhistorik (LeH) 39
4.5 Leistungshistorik Grundsicherung (LHG) 39
4.6 Arbeitsuchendenhistoriken (ASU/XASU) 40
4.6.1 ASU 40
4.6.2 XASU 41
4.7 Maßnahmeteilnahmehistoriken (MTH/XMTH) 42
4.7.1 MTH 42
4.7.2 XMTH 42
5 Variablenbeschreibung 44
5.1 Identifikatoren 44
5.1.1 Systemfreie Personennummer (persnr) 44
5.1.2 Systemfreie Betriebsnummer GAV-ADIAB (betnr_gavadiab) 44
5.2 Generierte technische Merkmale 45
5.2.1 Teil der GAV Stichprobe (gav) 45
5.2.2 Satzzähler pro Konto (spell) 45
5.2.3 Quelle des Satzes (quelle) 46
5.2.4 Jahr (jahr) 46
5.3 Gültigkeitszeitraum 46
5.3.1 Beginndatum Originalbeobachtung (begorig) 46
5.3.2 Enddatum Originalbeobachtung (endorig) 47
5.3.3 Beginndatum Episode (begepi) 48
5.3.4 Enddatum Episode (endepi) 48
5.4 Informationen zur Person 48
5.4.1 Geschlecht (frau) 48
5.4.2 Geburtsjahr (gebjahr) 48
5.4.3 Geburtsmonat (gebmon) 49
5.4.4 Staatsangehoerigkeit, vergroebert (deutsch) 49
5.4.5 Familienstand (famst) 50
5.4.6 Kinderzahl (kind) 50
5.4.7 Ausbildung (ausbildung) 51
5.4.8 Ausbildung (imputiert) (ausbildung_imp) 54
5.4.9 Schulausbildung (schule) 54
5.5 Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche 55
5.5.1 Tagesentgelt/ täglicher Leistungssatz (tentgelt) 55
5.5.2 Beruf – ausgeübte/letzte Tätigkeit (KldB 1988) (beruf) 57
5.5.3 Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_3) 58
5.5.4 Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_4) 59
5.5.5 Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau) 60
5.5.6 Teilzeit (teilzeit) 61
5.5.7 Stellung im Beruf und Arbeitszeit (stib) 62
5.5.8 Erwerbsstatus (erwstat) 62

5.5.9 Gleitzone (gleitz) 64
5.5.10 Leiharbeit (leih) 65
5.5.11 Befristung (befrist) 65
5.5.12 Abmeldegrund / Abgabegrund / Beendigungsgrund (grund) 65
5.5.13 Erwerbsstatus vor Arbeitsuche (estatvor) 67
5.5.14 Status nach Arbeitsuche (estatnach) 67
5.5.15 Integrationsprognose (ipo) 68
5.5.16 Art der Kündigung der letzten Tätigkeit (art_kuend) 69
5.5.17 Arbeitszeit des Stellengesuchs (arbzeit) 69
5.5.18 Restanspruch/geplante Dauer (restanspruch) 69
5.5.19 Trägerart (traeger) 70
5.5.20 Beginndatum der Arbeitslosigkeit (alo_beg) 70
5.5.21 Dauer der Arbeitslosigkeit (alo_dau) 71

5.6 Ortsangaben 71
5.6.1 Wohnort - Kreis (wo_kreis) 71
5.6.2 Wohnort - Bundesland (wo_bula) 72
5.6.3 Wohnort - Arbeitsagentur (wo_aa) 73
5.6.4 Wohnort - Regionaldirektion (wo_rd) 74

5.7 Betriebsmerkmale 74
5.7.1 WZ73 3-Steller (w73_3) 74
5.7.2 WZ73 aggregiert (w73_2_agg) 74
5.7.3 WZ93 5-Steller (w93_5) 75
5.7.4 WZ93 3-Steller (w93_3) 75
5.7.5 WZ03 5-Steller (w03_5) 76
5.7.6 WZ03 3-Steller (w03_3) 76
5.7.7 WZ08 5-Steller (w08_5) 77
5.7.8 WZ08 3-Steller (w08_3) 77
5.7.9 w73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen) 78
5.7.10 Art der Vervollständigung w73_3 (group_w73_3) 78
5.7.11 w93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen) 79
5.7.12 Art der Vervollständigung w93_3 (group_w93_3) 79
5.7.13 w08_3 vervollständigt durch Extrapolation/Imputation (w08_3_gen) 79
5.7.14 Art der Vervollständigung w08_3 (group_w08_3) 80
5.7.15 Jahr des 1. Auftretens der Betriebsnummer (grd_jahr) 80
5.7.16 Erstes Auftreten Betriebsnummer (grd_dat) 80
5.7.17 Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr) 81
5.7.18 Letztes Auftreten Betriebsnummer (lzt_dat) 81
5.7.19 Anzahl Beschäftigte gesamt (az_ges) 82
5.7.20 Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz) 82
5.7.21 Anzahl geringfügig Beschäftigte (az_gf) 82
5.7.22 Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw) 83
5.7.23 Arbeitsort - Kreis (ao_kreis) 83
5.7.24 Arbeitsort - Bundesland (ao_bula) 84

6 Variablenbeschreibung der GAV-spezifischen Merkmale 85

6.1 Informationen zur Person 85

6.1.1 Anzahl der Kinder (Kategorien) (GAV_kinder_cat) 85

6.1.2 Schulbildung (GAV_educ) 85

6.1.3 Universitaetsabschluss (GAV_uniabschl) 85

6.2 Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche 86

6.2.1 Beruf - aktuelle Taetigkeit ist ausbildungsgerecht (GAV_beruf_pass) 86

6.2.2 Beruf - Jahr des Abschlusses, aktueller Taetigkeit (GAV_beruf_jahr) 86

6.2.3 Beruf - erste erlernte Taetigkeit (KldB 1988), falls mehrere (GAV_beruf_e) 86

6.2.4 Beruf - erste erlernte Taetigkeit (KldB 2010), falls mehrere (GAV_beruf2010_e) 87

6.2.5 Beruf - Jahr des Abschlusses, erste Taetigkeit, falls mehrere (GAV_beruf_jahr_e) 87

6.2.6 Stellung im Beruf (GAV_stib) 87

6.2.7 Schichtarbeit in Vollzeit (GAV_schicht_vz) 88

6.2.8 Woechentliche Arbeitszeit (GAV_az) 88

6.2.9 Mehrbetriebsunternehmen (GAV_mehrfbet) 88

7 Literatur 89

8 Anhang 92

8.1 Auszählungen 92

8.2 Abkürzungsverzeichnis 92

Zusammenfassung

Dieser Datenreport beschreibt die Daten aus dem Datenspeicher Gesellschaftliches Arbeitsvermögen des Bundesarchivs zu Beschäftigung in der DDR im Jahr 1989, verknüpft mit den administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) (GAV-ADIAB) 1975 - 2019.

Abstract

This data report describes the Data from the Data Fund of Societal Work Power from the Federal Archive of Germany on employment in the GDR in 1989 linked with administrative data of the Institute for Employment Research (IAB) (GAV-ADIAB) 1975 - 2019.

Keywords

German administrative micro data, labour market data, data manual, Data Fund of Societal Work Power, DDR-Daten, GDR data

Danksagung

Für ihre Mitarbeit und Unterstützung danken wir den Kolleginnen und Kollegen des Forschungsdatenzentrums (FDZ) sowie des Geschäftsbereichs Daten- und IT-Management (DIM) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB). Außerdem danken wir Jeannette Madarász-Lebenhagen, Chris Berthold, Beate Friedrich, Petra Rauschenbach und Kerstin Risse vom Bundesarchiv für die Unterstützung bei diesem Projekt. Besonderer Dank gilt auch Alexandra Spitz-Oener, Jan Nimczik und Ole Monscheuer dafür, dass sie uns ihre Aufbereitungsprogramme zur Verfügung gestellt haben. Für den vorliegenden Datenreport wurden einzelne Passagen aus IAB-internen Datendokumentationen von IAB-DIM übernommen.

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

Die Arbeitgebermeldungen an die deutsche Sozialversicherung und Daten aus den internen Prozessen der Bundesagentur für Arbeit (BA) sind wichtige Quellen, um Arbeitsmarktbiografien von Personen in Deutschland abzubilden. Diese Informationen werden im Institut für Arbeitsmarkt- und Berufsforschung (IAB) der BA für Forschungszwecke zusammengeführt und zu den „Integrierten Erwerbsbiografien“ (IEB) aufbereitet. Die IEB bilden die Beschäftigung und Arbeitslosigkeit seit 1975 ab. Allerdings existieren vollständige Daten für Ostdeutschland dort erst ab 1992, was beispielsweise eine Analyse von den Effekten der Wiedervereinigung mit diesen Daten erheblich erschwert.

Um diese Lücke teilweise zu schließen, wurden, in einem ersten Schritt, Informationen über Personen, die 1989 in der ehemaligen Deutschen Demokratischen Republik (DDR) beschäftigt waren, aus dem Datenspeicher Gesellschaftliches Arbeitsvermögen (GAV) des Bundesarchivs mit ihren Arbeitsmarktbiografien aus den IEB zusammengespielt (siehe Liepmann und Müller, 2018). In einem zweiten Schritt wurden diese Daten um die Arbeitsmarktbiografien einer Vergleichsgruppe von Personen ergänzt, die 1989 in Westdeutschland beschäftigt waren. Darüber hinaus wurden die DDR-Daten aus dem GAV, so weit wie möglich, an die Struktur der IEB angepasst, sodass nun mit dem GAV-ADIAB 7519 erstmals ein Datensatz vorliegt, der die simultane Analyse von detaillierten individuellen Arbeitsmarktbiografien von Ost- und Westdeutschen ab einem Zeitpunkt *vor* der Wiedervereinigung bis hin zum Jahr 2019 ermöglicht.

Somit stellen Personen mit Beschäftigung im Jahr 1989 die Grundgesamtheit des GAV-ADIAB 7519. Deshalb gilt hinsichtlich seines Analysepotentials insbesondere zu beachten, dass mit dem GAV-ADIAB 7519 keine repräsentativen Analysen von Unterschieden zwischen Ost- und Westdeutschland zu einem Zeitpunkt *nach* der Wiedervereinigung möglich sind. Für den zuletzt gennannten Zweck sei beispielsweise auf die Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) verwiesen (siehe vom Berge et al, 2021). Der GAV-ADIAB 7519 hingegen dient der Analyse der langfristigen Entwicklung von Arbeitsmarktbiografien von Personen, die 1989 in der ehemaligen DDR oder der Bundesrepublik Deutschland gearbeitet haben. Der verknüpfte Datensatz ermöglicht somit nicht nur die Untersuchung von Änderungen im Erwerbsleben, wie Arbeitslosigkeit oder berufliche und regionale Mobilität, sondern kann außerdem genutzt werden, um bestehende Erkenntnisse zu den Konsequenzen der deutschen Wiedervereinigung für individuelle Arbeitsmarktbiografien zu erweitern.

Der GAV-ADIAB 7519 umfasst eine Zufallsstichprobe von 10% der Personen mit einem Eintrag in den GAV-Daten und deren Arbeitsmarktbiografien aus den IEB mittels Record-Linkage zugespielt werden konnten (für Details zur Zuspielung, siehe Kapitel 1.2.4). Zusätzlich enthält er eine 2,5-%-ige Zufallsstichprobe der Personen aus den IEB mit Beschäftigung im Jahr 1989 (für Details zum Stichprobendesign, siehe Kapitel 3.3). „Ostdeutsche 1989“ sind also in der Grundgesamtheit des GAV-ADIAB 7519 im Vergleich zu „Westdeutschen 1989“ vierfach überrepräsentiert. Durch die so erhöhte Zahl an Beobachtungen werden detailliertere Analysen mit den GAV-Daten ermöglicht (für Details zur Gewichtung, siehe Kapitel 3.4). Für alle selektierten Personen sind die vollständigen Daten aus den IEB bis maximal 2019 enthalten.

Die IEB enthalten Daten zu folgenden Zuständen:

- 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 arbeitsmarktpolitischer Maßnahme (erfasst ab 2000).

Diese aus unterschiedlichen Datenquellen stammenden Informationen werden in den IEB zusammengeführt und tagesgenau abgebildet.

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 den Maßnahmeteilnahmehistoriken (MTH und XMTH) erfasst wird.

Der GAV-ADIAB 7519 umfasst die Erwerbsverläufe von 1.113.908 Individuen, deren Erwerbsbiografien in insgesamt 58.157.609 Datenzeilen dokumentiert sind. Der vorliegende Datenreport beschreibt die Merkmale des GAV-ADIAB 7519, welcher weitgehend die Originaldaten, d.h. nicht vergröberte Daten, beinhaltet. Die Angaben aus dem GAV wurden dabei an die Struktur der IEB angepasst, wann immer dies sinnvoll möglich war. Einige Merkmale sind als besonders sensibel eingestuft und werden daher nur auf gesonderten Antrag weitergegeben (siehe Kapitel 1.2).

Der GAV-ADIAB 7519 steht über das Forschungsdatenzentrum (FDZ) der BA im IAB in schwach anonymisierter Form ausschließlich für Zwecke der nicht-kommerziellen Arbeitsmarkt- und Berufsforschung zur Verfügung. Dieser Datenreport beschreibt die verknüpften Daten des GAV-ADIAB 7519. Er gliedert sich wie folgt: Neben der Einleitung enthält Kapitel 1 unter anderem Informationen zum Datenzugang sowie eine Kurzbeschreibung der Daten, das Mengengerüst und eine Merkmalsliste. Eine Beschreibung der einzelnen Datenquellen findet sich in Kapitel 2. Die Datenaufbereitung sowie die Datenqualität werden in den Kapiteln 3 und 4 diskutiert, während die einzelnen Merkmale in den Kapiteln 5 und 6 dargestellt werden.

1.2 Datennutzung

1.2.1 Datenzugang

Die GAV-ADIAB-Daten in der schwach anonymisierten Version können ausschließlich im Rahmen von Gastaufenthalten am FDZ und anschließender Datenfernverarbeitung ausgewertet werden.

Für die Datennutzung ist zunächst ein Antrag beim FDZ einzureichen. Über die Genehmigung des Forschungsprojekts entscheidet das FDZ im Auftrag und ggf. unter Abstimmung mit dem

Bundesministerium für Arbeit und Soziales (BMAS). Nach erfolgter Genehmigung wird ein Nutzungsvertrag mit der wissenschaftlichen Einrichtung der Forschenden geschlossen. Details zur Beantragung des Datensatzes und Datenverarbeitungsmöglichkeiten können der FDZ-Website entnommen werden.

Sollte nur der GAV-Datensatz alleine, ohne die Verknüpfung mit den administrativen Daten des IAB, für Forschungsarbeiten benötigt werden, dann ist dieser beim Bundesarchiv zu beantragen (https://www.bundesarchiv.de/DE/Navigation/Benutzen/Hilfe/Anfragen-Bundesarchiv/anfragen-bundesarchiv.html).

1.2.2 Datenhaltung

Die GAV-ADIAB-Daten sind modular gegliedert und in mehreren Dateien abgelegt. Die Daten aus dem GAV zu Beschäftigten in der DDR im Jahr 1989 sind als Querschnitt in einem ersten Modul abgelegt. Diese GAV-Datei enthält die systemfreie Personennummer als Identifikator, persönliche Merkmale, Informationen zu Beschäftigung und dem beschäftigenden Betrieb, Merkmale zum Wohnort sowie technische Merkmale. Diese Merkmale sind teilweise mit den entsprechenden Merkmalen aus den administrativen Daten harmonisiert und teilweise GAV-spezifisch. Tabelle 1 gibt einen Überblick über die in der GAV-Datei enthaltenen Merkmale und deren Art der Harmonisierung. Details zu den einzelnen Merkmalen finden sich in den Kapiteln 5 und 6. Sämtliche Variablen- und Wertelabels sind sowohl in deutscher Sprache als auch in englischer Sprache enthalten.[1]

Tabelle 1: In der GAV-Datei enthaltene Variablen

Variable	**Art der Harmonisierung zwischen den GAV- und den administrativen Daten**
Harmonisiert – und somit sowohl in den GAV- als auch in den administrativen Daten enthalten	
Systemfreie Personennummer (persnr)	Identisch
Teil der GAV Stichprobe (gav)	Identisch
Quelle des Satzes (quelle)	Hinzugefügt als zusätzlicher Wert – Wert „9“ nur in GAV-Datei
Geschlecht (frau)	Identisch
Geburtsjahr (gebjahr)	Identisch
Geburtsmonat (gebmon)	Identisch (sensible Variable)
Staatsangehoerigkeit, vergroebert (deutsch)	Umgeschlüsselt und aggregiert
Familienstand (famst)	Hinzugefügt als zusätzliche Werte – Werte „111-114“ nur in GAV-Datei
Beruf – ausgeübte/letzte Tätigkeit (KldB 1988) (beruf)	Umgeschlüsselt
Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_3)	Umgeschlüsselt

[1] Mit den Stata-Befehlen *label language en* oder *label language de* kann auf englisch- bzw. deutschsprachige Labels umgestellt werden.

Variable	Art der Harmonisierung zwischen den GAV- und den administrativen Daten
Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau)	Umgeschlüsselt
Teilzeit (teilzeit)	Umgeschlüsselt
Wohnort - Kreis (wo_kreis)	Umgeschlüsselt (sensible Variable)
Wohnort - Bundesland (wo_bula)	Umgeschlüsselt
Arbeitsort - Kreis (ao_kreis)	Umgeschlüsselt (sensible Variable)
Arbeitsort - Bundesland (ao_bula)	Umgeschlüsselt
WZ73 aggregiert (w73_2_agg)	Umgeschlüsselt und aggregiert
GAV-spezifisch – nur in GAV-Datei enthalten	
Anzahl der Kinder (Kategorien) (GAV_kinder_cat)	GAV-spezifisch
Schulbildung (GAV_educ)	GAV-spezifisch
Universitaetsabschluss (GAV_uniabschl)	GAV-spezifisch
Beruf - aktuelle Taetigkeit ist ausbildungsgerecht (GAV_beruf_pass)	GAV-spezifisch
Beruf - Jahr des Abschlusses, aktueller Taetigkeit (GAV_beruf_jahr)	GAV-spezifisch
Beruf - erste erlernte Taetigkeit (KldB 1988), falls mehrere (GAV_beruf_e)	Umgeschlüsselt aber GAV-spezifisch
Beruf - erste erlernte Taetigkeit (KldB 2010), falls mehrere (GAV_beruf2010_e)	Umgeschlüsselt aber GAV-spezifisch
Beruf - Jahr des Abschlusses, erste Taetigkeit, falls mehrere (GAV_beruf_jahr_e)	GAV-spezifisch
Stellung im Beruf (GAV_stib)	GAV-spezifisch
Schichtarbeit in Vollzeit (GAV_schicht_vz)	GAV-spezifisch
Woechentliche Arbeitszeit (GAV_az)	GAV-spezifisch
Mehrbetriebsunternehmen (GAV_mehrfbet)	GAV-spezifisch

Ein zweites Modul, das im weiteren Verlauf Personendatei genannt wird, enthält die individuellen Arbeitsmarktbiografien aus den IEB zu allen im GAV-ADIAB 7519 enthaltenen Personen. Es umfasst Identifikatoren (systemfreie Personen- und Betriebsnummern), persönliche Merkmale, Informationen zu Beschäftigung, Leistungsbezug, Arbeitsuche und Teilnahme an arbeitsmarktpolitischen Maßnahmen, Merkmale zum Wohnort sowie generierte technische Merkmale.

Ein drittes Modul, die Basisbetriebsdatei, enthält die Betriebsnummer, das Jahr und Merkmale zu Arbeitsort und Wirtschaftszweigen sowie weitere Betriebsinformationen zum Stichtag 30. Juni. Für Ostdeutschland liegen die Betriebsdaten erst ab 1992 vor. Die Basisbetriebsdatei entstammt dem Betriebs-Historik-Panel (BHP) (siehe Ganzer et al., 2021). Für Merkmale aus den IEB und dem BHP sind sowohl deutsche als auch englische Variablen- und Wertelabel verfügbar.

Somit sind Betriebs- und Personenmerkmale sowie GAV- und IEB-Daten organisatorisch voneinander getrennt, was die Struktur der Daten deutlich macht und Speicherplatz spart (vgl. Abbildung 1). Die Verknüpfung der Betriebsdateien mit der Personendatei erfolgt über die Betriebsnummer und das Jahr des Datensatzes. Die Variablen in der Merkmalsliste (vgl. S. 17ff), die in der Spalte „BHP“ mit hohem (h) oder wechselnden (m) Füllgrad gekennzeichnet sind, sind

in der Basisbetriebsdatei enthalten. Zu beachten ist, dass das für die Verknüpfung der Module notwendige Merkmal „Betriebsnummer“ in allen Dateien außer der GAV-Datei vorhanden ist. Alle Dateinamen im GAV-ADIAB 7519 sind ebenfalls Abbildung 1 zu entnehmen.

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

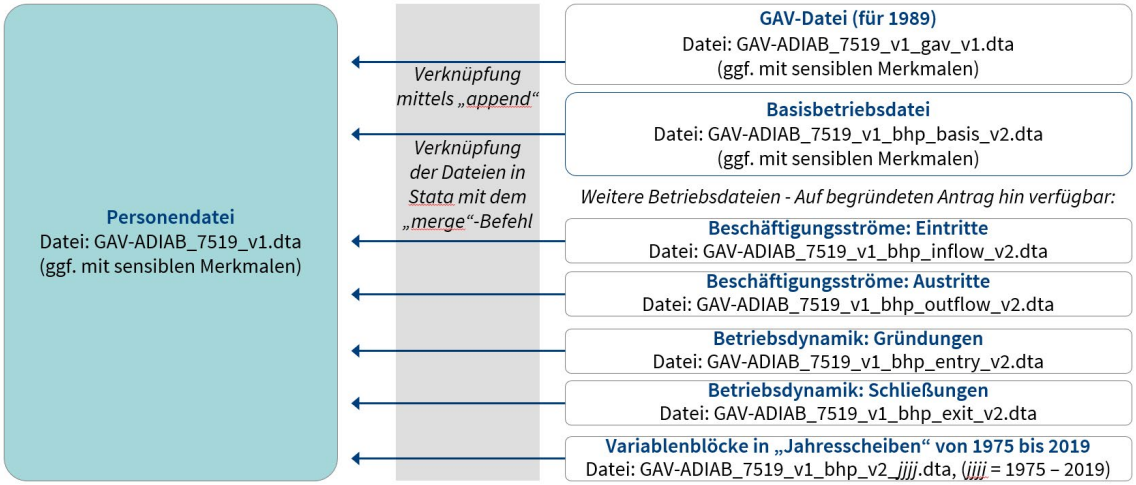
GAV-Datei:

- Geburtsmonat (gebmon)
- Wohnort - Kreis (wo_kreis)
- Arbeitsort - Kreis (ao_kreis)

Personendatei:

- Geburtsmonat (gebmon)
- Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_4)
- Wohnort - Arbeitsagentur (wo_aa)
- Wohnort - Kreis (wo_kreis)

Abbildung 1: Datenhaltung des GAV-ADIAB 7519



Basisbetriebsdatei:

- Arbeitsort - Kreis (ao_kreis)
- Wirtschaftszweig – Wirtschaftsunterklasse: WZ93 5-Steller (w93_5)
- Wirtschaftszweig – Wirtschaftsunterklasse: WZ03 5-Steller (w03_5)
- Wirtschaftszweig – Wirtschaftsunterklasse: WZ08 5-Steller (w08_5)
- Erstes Auftreten Betriebsnummer (grd_dat)
- Letztes Auftreten Betriebsnummer (lzt_dat)

Bei der Bereitstellung eines Datensatzes mit sensiblen Merkmalen werden die entsprechenden nicht-sensiblen Merkmale gelöscht, wenn diese von der oder dem Nutzenden selbst wieder generiert werden können. Beispielsweise wird die Variable Bundesland gelöscht, wenn die Variable Kreis zur Verfügung gestellt wird. Durch Abschneiden der letzten drei Stellen der Variable Kreis kann das Bundesland erzeugt werden.

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.

Für den GAV-ADIAB 7519 stehen Betriebsmerkmale aus dem BHP 1975-2019 zur Verfügung (siehe Ganzer et al., 2021). Dabei wird immer (d. h. auch ohne gesonderten Antrag) eine Basisbetriebsdatei bereitgestellt, die folgende Merkmale beinhaltet:

- Systemfreie Betriebsnummer GAV-ADIAB (betnr_gavadiab)
- Jahr (jahr)
- Arbeitsort - Bundesland (ao_bula)
- Klassifikation der Wirtschaftszweige 1973 – 3-Steller: WZ73 3-Steller (w73_3)
- Klassifikation der Wirtschaftszweige 1993 – 3-Steller: WZ93 3-Steller (w93_3)
- Klassifikation der Wirtschaftszweige 2003 – 3-Steller: WZ03 3-Steller (w03_3)
- Klassifikation der Wirtschaftszweige 2008 – 3-Steller: WZ08 3-Steller (w08_3)
- Zeitkonsistente Klassifikationen der Wirtschaftszweige (3-Steller) nach Eberle et al. (FDZ-Methodenreport 05/2011): (w73_3_gen, w93_3_gen, w08_3_gen)
- WZ73 aggregiert und harmonisiert: WZ73 aggregiert (w73_2_agg)
- Jahr des 1. Auftretens der Betriebsnummer (grd_jahr)
- Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr)
- Anzahl Beschäftigte gesamt (az_ges)
- Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz)
- Anzahl geringfügig Beschäftigte (az_gf)
- Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)

Daneben können auch folgende Variablenblöcke aus dem BHP beantragt werden:

- Struktur der Beschäftigten allgemein (az_ges az_f az_reg az_gf az_azubi az_atz az_vz 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_med te_p25 te_p75 te_imp_mw 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)

Darüber hinaus können auch folgende Erweiterungsmodule des BHP beantragt werden:

- Erweiterungsmodul – 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_temp_f 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)
- Erweiterungsmodul – Betriebsdynamik (eintritt besch inflow betnr_gavadiab_vor besch_vor status_vor austritt besch outflow betnr_gavadiab_nach besch_nach status_nach)

Für alle Merkmale, die nicht in der Basisbetriebsdatei enthalten sind, muss die Erforderlichkeit begründet werden. Dabei hat dies bei den sensiblen Merkmalen für jedes einzelne Merkmal, bei den Variablenblöcken pro Block und für jedes einzelne Erweiterungsmodul zu erfolgen.

Nähere Informationen zum BHP finden sich unter https://fdz.iab.de.

1.2.3 Dateinamen der administrativen Daten und der GAV-Daten

GAV-Datei

GAV-ADIAB_7519_v1_gav_v1.dta

Personendatei

GAV-ADIAB_7519_v1.dta

Basisbetriebsdatei

GAV-ADIAB_7519_v1_bhp_basis_v2.dta

Zusatzmodule BHP:

Bestandsdaten mit Variablenblöcken

GAV-ADIAB_7519_v1_bhp_v2_jjjj.dta, jjjj = 1975 - 2019

Beschäftigtenströme

GAV-ADIAB_7519_v1_bhp_inflow_v2.dta

GAV-ADIAB_7519_v1_bhp_outflow_v2.dta

Betriebsdynamik

GAV-ADIAB_7519_v1_bhp_entry_v2.dta

GAV-ADIAB_7519_v1_bhp_exit_v2.dta

1.2.4 Datenverknüpfen

Um alle Informationen des GAV-ADIAB 7519 nutzen zu können, müssen die einzelnen Datensätze zusammengespielt werden. Hierbei ist die Reihenfolge der Datenverknüpfung von Bedeutung. Als Ausgangspunkte sollte grundsätzlich immer zuerst die Personendatei geladen werden. Wenn außerdem Variablen aus den Betriebsdateien für die Analyse benötigt werden, dann sollte anschließend die Personendatei mit der Basis-Betriebsdatei verknüpft werden (in Stata mittels ‚merge'-Befehl, siehe Kasten 1). Anderenfalls werde nicht alle Werte für die Betriebsvariablen korrekt übernommen. Analog können auch weitere BHP-Variablen aus den Zuspielmodulen angespielt werden.

Kasten 1: Beispielcode Stata 16: Verknüpfung einer Personendatei mit der Basis-Betriebsdatei

```
use "${orig}/GAV-ADIAB_7519_v1", clear
gen jahr = year(begepi)
label variable jahr Jahr
sort betnr_gavadiab jahr
merge m:1 betnr_gavadiab jahr ///
using "${orig}/GAV-ADIAB_7519_v1_bhp_basis_v2"
```

Abschließend sollten dann die Daten aus der GAV-Datei als eine zusätzliche Datenzeile angefügt werden (in Stata mittels ‚append'-Befehl, siehe Kasten 2). Die GAV-Daten fügen sich somit als ergänzende zeitliche Beobachtung in die Datenstruktur der IEB ein, während die Betriebsmerkmale aus dem BHP als zusätzliche Variablen vorhanden sind. Da es sich bei den GAV-Daten um Querschnittsinformationen handelt, enthalten sie keine Angaben zu Beginn und Ende von Episoden. Diese können ergänzt werden, siehe Beispielcode in Kasten 2, welcher die Datumsangaben für die GAV-Daten auf den 31.12.2012 setzt. Danach können die Daten erneut sortiert werden, beispielsweise nach Personenidentifikator, Beginndatum der Episode und Jahr.

Kasten 2: Beispielcode Stata 16: Appenden der GAV-Daten 1989

```
append using "${orig}/GAV-ADIAB_7519_v1_gav_v1"
```

```
foreach var of varlist begorig endorig begepi endepi {
      replace `var' = 10957 if quelle == 9
}
replace jahr = 1989 if quelle == 9
sort persnr begepi
```

1.3 Steckbrief

Tabelle 2: Steckbrief

Kategorie	Beschreibung
Themen/ Merkmalsgruppen	**Gesellschaftliches Arbeitsvermögen (GAV):** Vom Staatssekretariat für Arbeit und Löhne (SAL) erfasste Meldungen von Arbeitgebern über arbeitsökonomische Merkmale der erfassungspflichtigen Mitarbeitenden **Beschäftigtenhistorik (BeH):** Jahres- und Abmeldungen zur Sozialversicherung für sozialversicherungspflichtig und geringfügig Beschäftigte **Leistungsempfängerhistorik (LeH):** Informationen zu Leistungsbezug nach SGB III für Bezieher von Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld **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) **Arbeitsuchendenhistorik (ASU):** Informationen zur Arbeitsuche, die in BA-Verfahren erfasst werden **Arbeitsuchendenhistorik 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 **Maßnahmeteilnahmehistorik (MTH):** Informationen zu Teilnahmen an Maßnahmen der aktiven Arbeitsmarktförderung (keine Maßnahmen der zugelassenen kommunalen Träger im SGB II) **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
Untersuchungseinheit	GAV: Erfassungspflichtige Beschäftigte in der DDR IEB: Sozialversicherungspflichtig Beschäftigte (ab 1999 auch geringfügig Beschäftigte), Leistungsempfangende, Arbeitsuchende, Maßnahmeteilnehmende BHP: Betriebe
Fallzahlen	1.113.908 Personen

Kategorie	Beschreibung
	43.771.891 Originalsätze 58.157.609 überschneidungsfreie Sätze (nach Episodensplitting)
Zeitraum	Der Zeitraum hängt von der Datenquelle ab. GAV: Querschnitt zum 31.12.1989 BeH: 01.01.1975 - 31.12.2019 (bis 2016: 36-Monats-Datei 2017 und 2018: 18-Monats-Datei, 2019: 6-Monats-Datei) LeH: 01.01.1975 - 31.12.2019 ASU: 01.01.1997 - 31.12.2019 LHG: 01.01.2005 - 31.12.2019 XASU: 01.01.2005 - 31.12.2019 MTH: 01.01.2000 - 31.12.2019 XMTH: 01.01.2005 - 31.12.2019 BHP: Querschnitt zum 30.06. eines jeden Jahres von 1975 - 2019
Zeitbezug	GAV: Stichtag 31.12.2019 IEB: Tagesgenau BHP: Stichtag 30.06 eines Jahres
Regionale Gliederung	Bundesländer, Kreise
Gebietsstand	GAV: Gebietsstandsbereinigt zum 31.12.2014 IEB: Gebietsstandsbereinigt zum 31.12.2019
Erhebungsdesign	Zufallsstichprobe von Personen mit Beschäftigung im Jahr 1989, geschichtet gezogen in Abhängigkeit des Arbeitsorts (10% falls DDR; 2,5% falls BRD)
Beteiligte Institutionen	GAV: Staatssekretariat für Arbeit und Löhne (SAL) der DDR, Bundesarchiv IEB: Sozialversicherungsträger, Bundesagentur für Arbeit, kommunale Träger
Frequenz der Datensammlung	GAV: Abgeschlossen IEB: Laufend BHP: Laufend
Dateiformat/ -größe	Stata; GAV-Datei: 22,2 MB; Personendatei: 3,7 GB; Basisbetriebsdatei: 587,1 MB
Dateiorganisation	Die Daten werden in drei Dateien abgelegt. Zwei enthalten personenbezogene und die Dritte betriebsbezogene Informationen. Auf begründeten Antrag werden weitere Dateien mit zusätzlichen Betriebsinformationen bereitgestellt.
Datenzugang	Gastaufenthalt oder Datenfernverarbeitung
Anonymisierungsgrad	Schwach anonymisiert
Sensible Merkmale	GAV: Geburtsmonat (gebmon), Wohnort – Kreis (wo_kreis), Arbeitsort - Kreis (ao_kreis) IEB: Geburtsmonat (gebmon), Berufsuntergruppe (beruf2010_4), Wohnort – Arbeitsagentur (wo_aa), Wohnort – Kreis (wo_kreis)

Kategorie	Beschreibung
	BHP: Arbeitsort - Kreis (ao_kreis), Wirtschaftsunterklasse (w93_5), Wirtschaftsunterklasse (w03_5), Wirtschaftsunterklasse (w08_5), Erstes Auftreten der Betriebsnummer (grd_dat), Letztes Auftreten der Betriebsnummer (lzt_dat)
Zitierung der Daten und Datendokumentation	Daten: „Die Datengrundlage dieses Beitrags bilden die schwach anonymisierten Daten aus dem Datenspeicher Gesellschaftliches Arbeitsvermögen verknüpft mit administrativen Daten des IAB (GAV-ADIAB) 1975-2019. Der Datenzugang erfolgte über einen Gastaufenthalt am Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ) und/oder mittels kontrollierter Datenfernverarbeitung beim FDZ." DOI: 10.5164/IAB.GAV-ADIAB7519.de.en.v1 Datendokumentation: Zimmermann, Florian; Umkehrer, Matthias; Ruf, Kevin; Müller, Dana; Seth, Stefan (2022): Datenspeicher Gesellschaftliches Arbeitsvermögen verknüpft mit administrativen Daten des IAB (GAV-ADIAB) 1975-2019. FDZ-Datenreport, 13/2022 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.2213.de.v1
Datensatzversion	Schwach anonymisierte Version der Daten aus dem Datenspeicher Gesellschaftliches Arbeitsvermögen verknüpft mit administrativen Daten des IAB (GAV-ADIAB) – Version 7519 v1; DOI: 10.5164/IAB.GAV-ADIAB7519.de.en.v1

1.4 Merkmalsliste

Die Merkmalsübersicht in Tabelle 4 listet die Variablennamen und die längeren Merkmalsbezeichnungen auf. Weiterhin bietet sie eine Übersicht darüber, ob und wie gut Merkmale in den einzelnen Datenquellen gefüllt sind. Die Variablen in der Merkmalsliste, die in der Spalte „BHP“ bzw. „GAV“ mit hohem (h) oder wechselnden (m) Füllgrad gekennzeichnet sind, sind in der Basisbetriebsdatei bzw. der GAV-Datei enthalten. Tabelle 3 macht die Bedeutung der Schattierungen deutlich, die in Tabelle 4 den Füllgrad pro Merkmal und Quelle angeben.

Beispiel: Das Merkmal Tagesentgelt bzw. täglicher Leistungssatz ist nur für BeH- und LeH-Sätze gefüllt. Die Sätze der anderen Datenquellen enthalten in dieser Variablen den Missing-Wert „.n“.

Tabelle 3: Füllgrad der Merkmale

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

Eine weitere Besonderheit liegt darin, dass einige Merkmale je nach Datenquelle unterschiedliche Inhalte haben. So enthält der Erwerbsstatus für BeH-Sätze die Personengruppe des Meldeverfahrens, für LeH-Sätze die Leistungsart, für LHG-Sätze den Status SGB II, für ASU- bzw. XASU-Sätze den Arbeitsuche-Status und für MTH- bzw. XMTH-Sätze die Maßnahmekategorie. Nicht bei jedem Merkmal sind diese Unterschiede direkt aus dem Variablennamen ersichtlich.

Tabelle 4: Merkmalsliste mit Füllgrad

Merkmalsliste	Seite	BeH	LeH	LHG	ASU	X ASU	MTH	X MTH	BHP	GAV
Identifikatoren	**44**									
Systemfreie Personennummer (persnr)	44	h	h	h	h	h	h	h	l	h
Systemfreie Betriebsnummer GAV-ADIAB (betnr_gavadiab)	44	h	l	l	l	l	l	l	h	l
Generierte technische Merkmale	**45**									
GAV Stichprobe (gav)	45	h	h	h	h	h	h	h	l	h
Satzzähler pro Konto (spell)	45	h	h	h	h	h	h	h	l	l
Quelle des Satzes (quelle)	46	h	h	h	h	h	h	h	l	h
Jahr (jahr)	46	l	l	l	l	l	l	l	h	l
Gültigkeitszeitraum	**46**									
Beginndatum Originalbeobachtung (begorig)	46	h	h	h	h	h	h	h	l	l
Enddatum Originalbeobachtung (endorig)	47	h	h	h	h	h	h	h	l	l
Beginndatum Episode (begepi)	48	h	h	h	h	h	h	h	l	l
Enddatum Episode (endepi)	48	h	h	h	h	h	h	h	l	l
Informationen zur Person	**48**									
Geschlecht (frau)	48	h	h	h	h	h	h	h	l	h
Geburtsjahr (gebjahr)	48	h	h	h	h	h	h	h	l	h
Geburtsmonat (gebmon)	49	h	h	h	h	h	h	h	l	h
Staatsangehörigkeit vergröbert (deutsch)	49	h	h	h	h	h	h	h	l	h
Familienstand (famst)	50	l	h	h	h	h	m	h	l	h
Kinderzahl (kind)	50	l	h	h	m	m	m	m	l	l
Ausbildung (ausbildung)	51	m	l	l	h	m	h	m	l	l

Merkmalsliste	Sei-te	BeH	LeH	LHG	ASU	X ASU	MTH	X MTH	BHP	GAV
Ausbildung (imputiert) (ausbildung_imp)	54	h	l	l	l	l	l	l	l	l
Schulausbildung (schule)	54	m	l	l	h	m	h	m	l	l
Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche	**55**									
Tagesentgelt/ täglicher Leistungssatz (tentgelt)	55	h	h	l	l	l	l	l	l	l
Beruf – ausgeübte/letzte Tätigkeit (KldB 1988) (beruf)	57	h	l	l	m	l	m	l	l	h
Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_3)	58	h	m	l	m	m	m	l	l	h
Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_4)	59	h	m	l	m	m	m	l	l	l
Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau)	60	h	m	l	m	m	m	l	l	h
Teilzeit (teilzeit)	61	h	l	l	l	l	l	l	l	h
Stellung im Beruf und Arbeitszeit (stib)	62	h	l	l	l	l	l	l	l	l
Erwerbsstatus (erwstat)	62	h	h	h	h	h	h	h	l	l
Gleitzone (gleitz)	64	h	l	l	l	l	l	l	l	l
Leiharbeit (leih)	65	h	l	l	l	l	l	l	l	l
Befristung (befrist)	65	h	l	l	l	l	l	l	l	l
Abmeldegrund / Abgabegrund / Beendigungsgrund (grund)	65	h	h	m	h	m	l	l	l	l
Erwerbsstatus vor Arbeitsuche (estatvor)	67	l	l	l	h	m	l	l	l	l
Status nach Arbeitsuche (estatnach)	67	l	l	l	m	l	l	l	l	l
Integrationsprognose (ipo)	68	l	l	l	m	l	m	l	l	l
Art der Kündigung der letzten Tätigkeit (art_kuend)	69	l	l	l	m	l	m	l	l	l
Arbeitszeit des Stellengesuchs (arbzeit)	69	l	l	l	m	l	m	l	l	l
Restanspruch/geplante Dauer (restanspruch)	69	l	h	l	l	l	h	l	l	l
Trägerart (traeger)	70	l	l	h	h	h	h	h	l	l
Beginndatum der Arbeitslosigkeit (alo_beg)	70	m	m	m	m	m	m	m	l	l
Dauer der Arbeitslosigkeit (alo_dau)	71	h	h	h	h	h	h	h	l	l
Ortsangaben	**71**									
Wohnort - Kreis (wo_kreis)	71	h	h	h	h	h	h	h	l	h

Merkmalsliste	Sei-te	BeH	LeH	LHG	ASU	X ASU	MTH	X MTH	BHP	GAV
Wohnort - Bundesland (wo_bula)	72	h	h	h	h	h	h	h	l	h
Wohnort - Arbeitsagentur (wo_aa)	73	h	h	h	h	h	h	h	l	l
Wohnort - Regionaldirektion (wo_rd)	74	h	h	h	h	h	h	h	l	l
Betriebsmerkmale	**74**									
WZ73 3-Steller (w73_3)	74	l	l	l	l	l	l	l	h	l
WZ73 aggregiert (w73_2_agg)	74	l	l	l	l	l	l	l	h	h
WZ93 5-Steller (w93_5)	75	l	l	l	l	l	l	l	h	l
WZ93 3-Steller (w93_3)	75	l	l	l	l	l	l	l	h	l
WZ03 5-Steller (w03_5)	76	l	l	l	l	l	l	l	h	l
WZ03 3-Steller (w03_3)	76	l	l	l	l	l	l	l	h	l
WZ08 5-Steller (w08_5)	77	l	l	l	l	l	l	l	h	l
WZ08 3-Steller (w08_3)	77	l	l	l	l	l	l	l	h	l
w73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen)	78	l	l	l	l	l	l	l	h	l
Art der Vervollständigung w73_3 (group_w73_3)	78	l	l	l	l	l	l	l	h	l
w93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen)	79	l	l	l	l	l	l	l	h	l
Art der Vervollständigung w93_3 (group_w93_3)	79	l	l	l	l	l	l	l	h	l
w08_3 vervollständigt durch Extrapolation/Imputation (w08_3_gen)	79	l	l	l	l	l	l	l	h	l
Art der Vervollständigung w08_3 (group_w08_3)	80	l	l	l	l	l	l	l	h	l
Jahr des 1. Auftretens der Betriebsnummer (grd_jahr)	80	l	l	l	l	l	l	l	h	l
Erstes Auftreten Betriebsnummer (grd_dat)	80	l	l	l	l	l	l	l	h	l
Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr)	81	l	l	l	l	l	l	l	h	l
Letztes Auftreten Betriebsnummer (lzt_dat)	81	l	l	l	l	l	l	l	h	l
Anzahl Beschäftigte gesamt (az_ges)	82	l	l	l	l	l	l	l	h	l
Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz)	82	l	l	l	l	l	l	l	h	l

Merkmalsliste	Sei-te	BeH	LeH	LHG	ASU	X ASU	MTH	X MTH	BHP	GAV
Anzahl geringfügig Beschäftigte (az_gf)	82	l	l	l	l	l	l	l	h	l
Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)	83	l	l	l	l	l	l	l	h	l
Arbeitsort - Kreis (ao_kreis)	83	l	l	l	l	l	l	l	h	h
Arbeitsort - Bundesland (ao_bula)	84	l	l	l	l	l	l	l	h	h
Variablenbeschreibung der GAV-spezifischen Merkmale										
Anzahl der Kinder (Kategorien) (GAV_kinder_cat)	85	l	l	l	l	l	l	l	l	h
Schulausbildung (GAV_educ)	85	l	l	l	l	l	l	l	l	h
Universitätsabschluss (GAV_uniabschl)	85	l	l	l	l	l	l	l	l	h
Beruf - aktuelle Tätigkeit ist ausbildungsgerecht (GAV_beruf_pass)	86	l	l	l	l	l	l	l	l	h
Beruf – Jahr des Abschlusses, aktueller Beruf (GAV_beruf_jahr)	86	l	l	l	l	l	l	l	l	h
Beruf - erste erlernte Tätigkeit (KldB 1988), falls mehrere (GAV_beruf_e)	86	l	l	l	l	l	l	l	l	m
Berufsgruppe - erste erlernte Taetigkeit (KldB 2010), falls mehrere (GAV_beruf2010_e)	87	l	l	l	l	l	l	l	l	m
Beruf - Abschlussjahr der ersten erlernten Tätigkeit (GAV_beruf_jahr_e)	87	l	l	l	l	l	l	l	l	m
Stellung im Beruf (GAV_stib)	87	l	l	l	l	l	l	l	l	h
Schicht Vollzeit (GAV_schicht_vz)	88	l	l	l	l	l	l	l	l	h
Wöchentliche Arbeitszeit (GAV_az)	88	l	l	l	l	l	l	l	l	h
Mehrbetriebsunternehmen (GAV_mehrfbet)	88	l	l	l	l	l	l	l	l	h

1.5 Mengengerüst

Tabelle 5: Mengengerüst

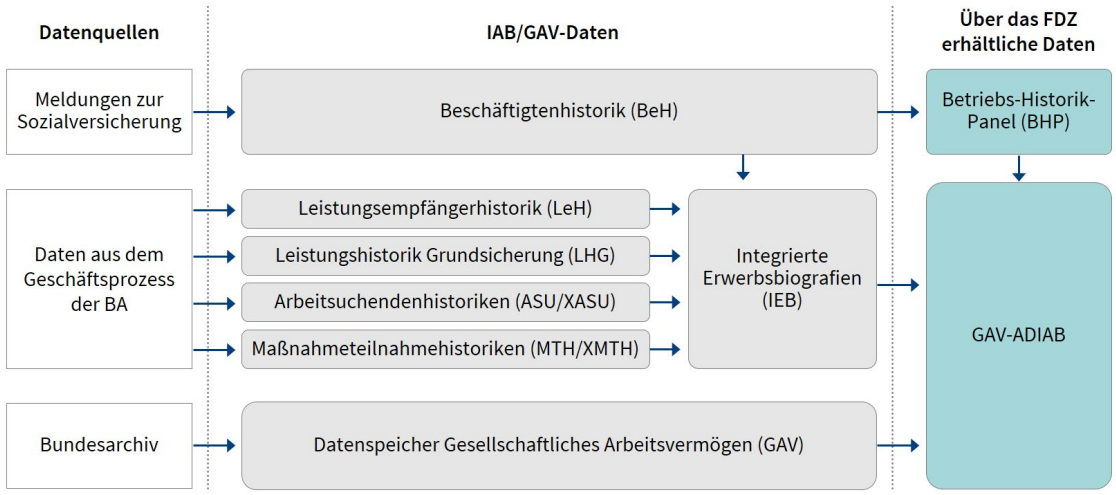
Fallzahlen	**vor Episodensplitting**	**nach Episodensplitting**
BeH	31.114.586	35.457.212
LeH	4.599.276	7.027.810
LHG	940.652	2.977.841
ASU	5.228.970	9.709.842
XASU	391.492	694.877
MTH	888.231	1.602.538
XMTH	70.058	148.863
GAV	538.626	538.626

Fallzahlen	vor Episodensplitting	nach Episodensplitting
Sätze insgesamt	43.771.891	58.157.609
Personen	1.113.908	1.113.908

2 Datenquellen

Die administrativen Personendaten werden aus den Integrierten Erwerbsbiografien (IEB) des IAB gezogen. Diese vereinen Daten aus 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. Außerdem werden Daten aus dem Datenspeicher Gesellschaftliches Arbeitsvermögen (GAV) für die Personen, die im Jahr 1989 in der DDR beschäftigt waren, zugespielt. Abbildung 2 veranschaulicht die Datenflüsse, die zum GAV-ADIAB 7519 führen und stellt deren Beziehung zu anderen Datenprodukten des FDZ dar.

Abbildung 2: Mengengerüst



2.1 GAV-Daten

Die Daten aus dem Datenspeicher Gesellschaftliches Arbeitsvermögen (GAV) wurden zur Erfassung und Auswertung von erfassungspflichtigen Beschäftigten vom Staatssekretariat für Löhne und Arbeit (SAL) erhoben und umfassen mit über 7 Millionen Personen ungefähr 72% der werktätigen Bevölkerung der DDR. Von dieser Erfassungspflicht ausgeschlossen waren insbesondere Mitarbeitende des Sicherheitsapparats, Mitarbeitende der Volksbildung, Leitungspersonal und Beschäftigte der Staatsorgane. Erfassungspflichtig hingegen waren u.a. Arbeiterinnen bzw. Arbeiter und Angestellte mit einem unbefristeten oder befristeten Arbeitsvertrag, Mitglieder von Produktionsgenossenschaften und Rechtsanwaltskollegien, noch erwerbstätige Ruheständler

sowie Wehrpflichtige und Zivildienstleistende (siehe Dietz und Rudolph 1990, Rathje 1996, Gebauer et al. 2004).

Die GAV-Daten waren nicht öffentlich zugänglich, sondern sollten das SAL bei der Planung und Kontrolle von Arbeitskräftebewegungen unterstützen (Salheiser, 2006). Sie wurden dezentral mittels einer Meldepflicht erhoben und setzten sich ursprünglich aus monatlichen Meldungen der Personalleitungen der Betriebe über die arbeitsökonomischen Merkmale aller Beschäftigten zusammen, die am SAL zusammenliefen. Von dem gesamten Datensatz ist allerdings nur noch der Stichtagsdatenbestand vom 31. Dezember 1989 vorhanden. Deshalb sind mit den GAV-Daten an sich nur Querschnittsanalysen möglich (Rathje, 1996). Weitere Informationen zu den GAV-Daten können auch dem Methodenreport von Liepmann und Müller (2018) entnommen werden.

Die im GAV erhobenen Merkmale lassen sich grob in vier Kategorien untergliedern:

1. Demographie (z.B. Alter, Geschlecht, Wohnort, Anzahl Kinder unter 14 Jahre, Familienstand, Staatsangehörigkeit)
2. Qualifikation (z.B. Schulbildung, Ausbildung, Hochschulabschluss)
3. Beschäftigung (z.B. Status, Arbeitsort, Tätigkeit, Arbeitszeit, Beruf)
4. Arbeitgeber (z.B. Unternehmenstyp, Branche)

Einkommensinformationen wurden ursprünglich erhoben, aus Datenschutzgründen im Jahr 1991 aber wieder gelöscht. Eine Übersicht über die Variablen, die im GAV-ADIAB 7519 enthalten sind, bieten die Kapitel 5 und 6.

2.2 Administrative Daten aus den IEB

2.2.1 Beschäftigtenhistorik (BeH)

Die Quelle für Informationen über Beschäftigung ist die Beschäftigtenhistorik (BeH) des IAB. 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 und Cramer, 1988). Es verlangt von den Arbeitgebern, mindestens einmal jährlich Meldungen für alle sozialversicherungspflichtig beschäftigten Arbeitnehmerinnen und Arbeitnehmer an die zuständigen Sozialversicherungsträger abzugeben. Die BeH erfasst alle Arbeiterinnen und Arbeiter und Angestellten sowie alle Auszubildenden, soweit sie nicht von der Sozialversicherungspflicht befreit sind. Dies bedeutet, dass verbeamtete und selbstständige Tätigkeiten sowie Studierende[2] (vgl. Cramer, 1985) grundsätzlich nicht in der BeH erfasst werden. Seit der Änderung des Meldeverfahrens zum 01.01.1999 werden auch geringfügig Beschäftigte und mithelfende Familienangehörige erfasst (enthalten erst ab 01.04.1999). Die Daten werden von den Krankenkassen aufgenommen, von der Bundesagentur für Arbeit gesammelt und aufbereitet und anschließend vom IAB in dessen Historikdatei integriert.

Die Beschäftigungsepisoden in den administrativen Personendaten werden durch Betriebsdaten (Basisbetriebsdatei und BHP-Erweiterungsmodule) ergänzt. Diese werden dem Betriebs-Historik-Panel (BHP) entnommen, das ebenso auf der BeH basiert. Die Basisbetriebsdatei und BHP-Erweiterungsmodule, die mit den GAV-ADIAB-Daten bereitgestellt werden, umfassen nur Angaben

[2] Studierende können dennoch in der BeH erscheinen, wenn sie parallel zum Studium z.B. geringfügig beschäftigt waren.

zu Betrieben, für die mindestens eine beschäftigte Person in der Stichprobe des GAV-ADIAB 7519 enthalten ist. Des Weiteren sind die Betriebsdaten auf die Jahre begrenzt, in denen die beschäftigten Personen in der Stichprobe 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.[3]

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).

2.2.2 Leistungsempfängerhistorik (LeH)

Die Leistungsempfängerhistorik (LeH) erfasst Zeiträume, in denen Personen Lohnersatzleistungen der Bundesagentur für Arbeit beziehen (Rechtskreis SGB III). Die Leistungen beinhalten Arbeitslosengeld, Arbeitslosenhilfe, Unterhaltsgeld und von der Bundesagentur für Arbeit übernommene Beiträge zur privaten Kranken- bzw. Pflegeversicherung während des Leistungsbezugs. 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 erfasst. Der linke Datenrand der LeH ist der 01.01.1975.

2.2.3 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 innerhalb von Bedarfsgemeinschaften ist im GAV-ADIAB 7519 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 der GAV-ADIAB 7519 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:

[3] Extrembeispiel: Es liegen Beschäftigungsmeldungen vom 1.1. bis zum 30.05.2006 vor, der Betrieb meldet jedoch Konkurs an, weshalb alle Beschäftigten entlassen werden. Dann gibt es im BHP zu diesem Betrieb für 2006 keine Informationen.

- Arbeitsgemeinschaften (ARGEn, 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 Kommune[4],
- 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 ARGEn und ab 2011 bei gemeinsamen Einrichtungen Verwendung.[5] Zugelassene kommunale Träger verwenden verschiedene eigene IT-Verfahren und übermitteln ihre Daten mittels des Standards XSozial-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 4).

2.2.4 Arbeitsuchendenhistoriken (ASU/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 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.

2.2.5 Maßnahmeteilnahmehistoriken (MTH/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.

[4] 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 Eingliederungs-leistungen (SGB III und SGB II) sowie spezifische Leistungen ohne die oben genannten flankierenden Leistungen erbracht.

[5] 2014 wurde A2LL sukzessive durch ALLEGRO als neues IT-Verfahren für die Grundsicherung im SGB II in gemeinsamen Einrichtungen ersetzt.

3 Datenaufbereitung und Stichprobenziehung

3.1 GAV-Daten

3.1.1 Bereinigung und Aufbereitung

Ein zentrales Ziel der Datenaufbereitung für den GAV-ADIAB 7519 war es, die GAV-Daten möglichst nahe an die Daten der IEB anzupassen. Deshalb wurden einige Personengruppen aus den GAV-Daten entfernt. In den GAV-Daten grundsätzlich enthaltene Personengruppen, die aus dem GAV-ADIAB 7519 entfernt wurden, sind:

- Rentner im Arbeitsprozess
- Arbeiter und Angestellte in einem zweiten Arbeitsverhältnis (doppelt erfasst)
- Werktätige, die den Grundwehrdienst oder Reservistendienst ableisten
- ausländische Werktätige mit Wohnsitz in der DDR (ca. 22.000 der insgesamt ca. 93.000 ausländischen Beschäftigten)
- Selbständige und deren Angestellte
- Handwerker, sofern ihre Betriebe keiner Produktionsgenossenschaft angehören
- Pädagogische Fachkräfte im Bereich Schulen und Kinderbetreuung

Anschließend wurden folgende Bereinigungen durchgeführt:

- Zuspielung der Wirtschaftszweiginformation aus dem statistischen Betriebsregister
- Umschlüsselung und Harmonisierung der Variablen deutsch, famst, beruf, beruf2010_3, niveau, teilzeit, ao_kreis, wo_kreis, w73_2_agg, GAV_beruf2010_e und GAV_beruf_e (siehe Tabelle 1 für eine Übersicht und die entsprechenden Variablenbeschreibung in den Kapiteln 5 und 6 für Details)
- In seltenen Fällen: Bereinigung unvollständiger Informationen

Weitergehende Bereinigungen, wie z.B. das Ergänzen von fehlenden Werten oder Korrekturen scheinbar nicht plausibler Werte wurden nicht vorgenommen. Derartige Entscheidungen sind den Nutzerinnen und Nutzern überlassen.

3.1.2 Verknüpfung der Personendaten

In einem vorangegangenen Schritt wurden die GAV-Daten einer Person mittels Record-Linkage Techniken mit ihren Daten aus den IEB verknüpft. Eine detaillierte Beschreibung dieses Vorgehens findet sich in Liepmann und Müller (2018), sowie technische Details in Antoni (2018). Dieser Abschnitt beschränkt sich auf einen groben Überblick.

Die Verknüpfung der Datenquellen GAV und IEB erfolgte auf Grundlage von Nachnamen, Vornamen, Geburtsdatum und Geschlecht. Diese Identifikatoren wurden ausschließlich zum Zweck und nur für die Dauer der Verknüpfung angespielt. Im GAV-ADIAB 7519 sind lediglich systemfreie Nummern für Personen oder Betriebe als Identifikatoren enthalten.

Aufgrund ihrer Komplexität wurden die Daten der IEB vor der Zuspielung auf Personen eingeschränkt,

- die zwischen 1929 und 1975 geboren und somit 1989 zwischen 14 und 60 Jahre alt waren,
- für die in den IEB-Daten mindestens eine Episode zwischen 1990 und 1996 in Ost- oder Westdeutschland verzeichnet ist und
- für die vor 1990 keine Episoden in Westdeutschland erfasst wurden.

Ursprünglich konnten knapp 77% der Personen in den GAV-Daten mit ihren Arbeitsmarktbiographien aus den IEB verlinkt werden. Dies entspricht in etwa einem Anteil von 72% der Arbeitskräfte in der DDR im Jahr 1989. Für den GAV-ADIAB 7519 wurde allerdings eine aktuellere Version der IEB verwendet als bei der Durchführung der Verknüpfung. Aufgrund der hierzu notwendigen Umschlüsselungen ging zusätzlich ein geringer Anteil (weniger als 0,5%) der verlinkten Personen verloren. Insgesamt kann die Verknüpfungsqualität als sehr gut beurteilt werden. Für Personen, die im Jahr 1989 älter als 60 waren, und Frauen, die im Jahr 1989 nicht verheiratet waren, konnte aber relativ gesehen häufiger keine Zuspielung erreicht werden als für andere Gruppen (siehe Kapitel 4.1).

3.2 Administrative Daten aus den IEB

3.2.1 Bereinigungen und Aufbereitungen

Bevor die Daten aus den in Kapitel 2.2 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.
- Datensätze, bei denen das Enddatum kleiner als das Beginndatum ist, werden gelöscht.
- Widersprüchliche Angaben zu Geschlecht oder Geburtsdatum innerhalb eines Kontos werden bereinigt.
- Datensätze ohne Angabe von Geburtsdatum und Geschlecht nach Bereinigung werden gelöscht.

Weitergehende Bereinigungen, wie z.B. das Ergänzen von vermutlich fehlenden Meldungen, Streikbereinigungen, etc., erfolgen nicht.

3.2.1.1 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 31.12.2019.

3.2.1.2 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 Gebietsstandsbereinigung wird analog zur BeH durchgeführt.

3.2.1.3 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 3.2.2),
  - 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 4.2.2) wurden soweit möglich bereinigt.
- Eine Gebietsstandsbereinigung wird analog zur BeH durchgeführt.

3.2.1.4 Arbeitsuchendenhistoriken 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 arbeitsuchend zu arbeitslos) erfolgt. Dies gilt auch, wenn sich die Trägerart (Agentur für Arbeit, ARGE, gemeinsame Einrichtung, zugelassener kommunaler Träger, getrennte Aufgabenwahrnehmung) ändert.
- In der XASU werden überschneidungsfreie Zeiträume bezogen auf eine einzelne Person abgebildet. Bei Änderungen folgender Merkmale wird für die XASU jeweils ein neuer Datensatz gebildet:
  - Wechsel des Arbeitsuche-Status
  - Wechsel der Verfügbarkeit
  - Wechsel des SGB-II-Trägers
  - Wechsel des Wohnortes
- Eine Gebietsstandsbereinigung wird analog zur BeH durchgeführt.

3.2.1.5 Maßnahmeteilnahmehistoriken (MTH)

- Sätze, die mehr als ein Jahr nach Maßnahmeende erstellt wurden, werden gelöscht, falls ein anderer Satz existiert, der innerhalb eines Jahres nach Abschluss der Maßnahme erstellt wurde.
- Es wird lediglich der aktuellste 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 Gebietsstandsbereinigung wird analog zur BeH durchgeführt.

3.2.1.6 Maßnahmeteilnahmehistoriken aus XSozial (XMTH)

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

der eingeführten Förder-ID 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 Teilnehmendem teilweise oder vollständig zeitlich überlappen. Echte Dubletten werden identifiziert und aussortiert, da sie mit hoher Wahrscheinlichkeit technische Dubletten sind. Versteckte Dubletten sind ebenfalls identifizierbar und werden aussortiert. Dabei gilt im Prinzip immer die Regel, dass der aktuellste Informationsstand behalten wird.

- Daneben werden Meldungen mit einem Beginndatum vor 2005 ausgeschlossen.
- Überlappende und unmittelbar aneinander angrenzende Meldungen der gleichen 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 Gebietsstandsbereinigung wird analog zur BeH durchgeführt.

3.2.2 SGB-II-Anonymisierung

Um das Deanonymisierungsrisiko zu verringern, wird im GAV-ADIAB 7519 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 aber an diesem Tag neu beginnen.

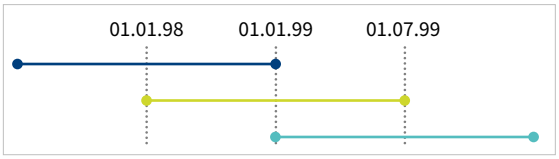
Um eine indirekte Bestimmung des exakten Geburtstags zu verhindern, wird daher das folgende Verfahren angewendet. Sätze, die am 18., 25. oder 65. Geburtstag geteilt wurden, werden zu einem einzelnen Satz zusammengefügt. Hierdurch geht außer dem exakten Geburtstag keine weitere Information verloren. Bei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

3.2.3 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).

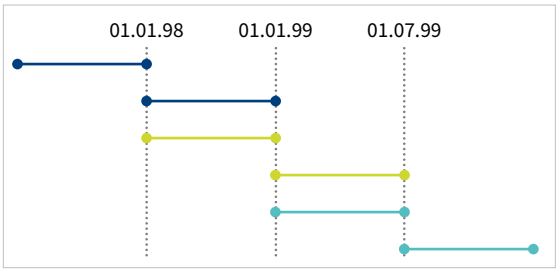
Die ursprünglichen Datumsvariablen zum Beginn und Ende des Originalsatzes (begorig und endorig) bleiben erhalten, die Variablen Beginn und Ende der Episode (begepi und endepi) kennzeichnen Anfang und Ende der gesplitteten Episoden. Ob Sätze gesplittet wurden, kann an einem Vergleich des ursprünglichen 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



persnr	*begorig*	*endorig*	*level*
4711	01.01.98	31.12.98	-
4711	01.10.98	30.06.99	-
4711	01.01.99	31.12.99	-

Episodensplitting
Aufteilung der Zeitintervalle in sich nicht überschneidende Intervalle



persnr	*begepi*	*endepi*	*level*
4711	01.01.98	30.09.98	0
4711	01.10.98	31.12.98	0
4711	01.10.98	31.12.98	1
4711	01.01.99	30.06.99	0
4711	01.01.99	30.06.99	1
4711	01.07.99	31.12.99	0

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

Kasten 3: Beispielcode zum Erstellen weiterer Satzzähler in Stata 16

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

3.3 Stichprobendesign und Analysepotenzial

Die IEB bilden die Beschäftigung und Arbeitslosigkeit seit 1975 ab. Allerdings existieren vollständige Daten für Ostdeutschland dort erst ab 1992. Um diese Lücke zumindest teilweise zu schließen, wurden die Personeninformationen aus den GAV-Daten von Beschäftigten in der DDR im Jahr 1989 mit ihren Arbeitsmarktbiographien aus den IEB verknüpft (siehe Kapitel 3.1.2). Außerdem enthält der GAV-ADIAB 7519 die Arbeitsmarktbiographien aus den IEB von Personen, die im Jahr 1989 in der Bundesrepublik Deutschland sozialversicherungspflichtig beschäftigt waren. Diese zweite Personengruppe soll als Vergleichsgruppe dienen und somit die simultane

Analyse von detaillierten individuellen Arbeitsmarktbiografien von Ost- und Westdeutschen ab einem Zeitpunkt *vor* der Wiedervereinigung bis hin zum Jahr 2019 ermöglichen. Die Vergleichsgruppe umfasst Personen,

- die zwischen 1929 und 1975 geboren und somit 1989 zwischen 14 und 60 Jahre alt waren,
- für die in den IEB-Daten mindestens eine Episode zwischen 1990 und 1996 in Ost- oder Westdeutschland verzeichnet ist und
- die im Jahr 1989 in den IEB als beschäftigt registriert waren.

Somit stellen Personen mit Beschäftigung im Jahr 1989 die Grundgesamtheit des GAV-ADIAB 7519. Deshalb gilt hinsichtlich seines Analysepotentials insbesondere zu beachten, dass mit dem GAV-ADIAB 7519 *keine* repräsentativen Analysen von Unterschieden zwischen Ost- und Westdeutschland zu einem Zeitpunkt *nach* der Wiedervereinigung möglich sind. Für den zuletzt gennannten Zweck sei beispielsweise auf die Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) verwiesen (siehe vom Berge et al, 2021). Der GAV-ADIAB 7519 hingegen dient der Analyse der langfristigen Entwicklung von Arbeitsmarktbiografien von Personen, die 1989 in der ehemaligen DDR oder der Bundesrepublik Deutschland gearbeitet haben. Der verknüpfte Datensatz ermöglicht somit nicht nur die Analyse von Phänomenen wie Arbeitslosigkeit, berufliche Mobilität und regionale Mobilität, sondern kann außerdem genutzt werden, um das bestehende Wissen über die Konsequenzen der deutschen Wiedervereinigung für individuelle Arbeitsmarktbiografien zu erweitern.

Der GAV-ADIAB 7519 umfasst eine Zufallsstichprobe von 10% der Personen aus dem GAV mit Verknüpfung zu den IEB. Außerdem enthält er eine Zufallsstichprobe von 2,5% der Personen aus der Vergleichsgruppe.

3.4 Gewichtung

Personen, die im Jahr 1989 entweder in der DDR oder der BRD beschäftigt waren, bilden die Grundgesamtheit des GAV-ADIAB 7519. Da es sich jeweils um reine Zufallsstichproben handelt, sind die Daten repräsentativ für diese Gruppen. Allerdings wurden Beschäftigte in der DDR viermal so häufig gezogen wie Beschäftigte in der BRD, um die Fallzahl in den GAV-Daten zu erhöhen. Folglich wird eine Gewichtung notwendig, sobald diese beiden Gruppen von Beschäftigten gemeinsam ausgewertet werden.

Im Fall von gemeinsamen Auswertungen müssen Personen mit Beschäftigung in der DDR im Jahr 1989 um den Faktor „4“ heruntergewichtet werden, um Repräsentativität der Ergebnisse zu erreichen. Alternativ können auch Personen mit Beschäftigung in der Bundesrepublik Deutschland um den Faktor „4“ hochgewichtet werden. Der folgende Code illustriert, wie die Gewichtung in Stata im verknüpften Datensatz umgesetzt werden könnte:

Kasten 4: Beispielcode zur Erstellung eines Gewichtungsfaktors in Stata 16

```
* Generiere Gewicht (ausgehend von den gemäß Kasten 1 und Kasten 2 zusammengespielten Daten)

generate weight = 1 if gav == 1
```

```
replace  weight = 4 if gav == 0
* Beispiel: Anteil Personen in GAV-Stichprobe
sort persnr begepi
tabulate gav [fweight=weight] if persnr != persnr[_n-1]
```

3.5 Fehlende Werte

In den Daten gilt die folgende Codierung von fehlenden Werten:

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

4 Datenqualität und Probleme

4.1 GAV

Grundsätzlich genügt die Qualität der GAV-Daten hohen Ansprüchen. Insbesondere wurden die von den Betrieben gemeldeten Informationen bei Ihrer Erfassung auf ihre Richtigkeit und Plausibilität hin überprüft und mussten gegebenenfalls korrigiert werden (Rathje, 1996). Allerdings fand diese Korrektur in einigen wenigen Fällen nicht statt, was fehlende Informationen zur Folge hatte (Dietz und Rudolph, 1990). Wie in Kapitel 2.1 erörtert sind einige Personengruppen in den GAV-Daten nicht enthalten. Dies betrifft insbesondere den so genannte "Sektor X", der ein integraler Bestandteil des DDR-Regimes war. Insgesamt gab es im Jahr 1989 in der DDR 9,75 Millionen Erwerbspersonen (Statistisches Bundesamt, 1994). Die mit den IEB verknüpfbaren GAV-Daten umfassen somit etwa 72% der ostdeutschen Erwerbsbevölkerung.

Hinsichtlich der Verknüpfbarkeit der GAV-Daten mit den IEB ist die Verknüpfungsqualität mit einer durchschnittlichen Trefferquote von knapp 77% als sehr gut einzuschätzen (siehe Antoni, 2018). Allerdings sind zwei Personengruppen in der Grundgesamtheit des GAV hervorzuheben, die eine auffällig schlechte Verknüpfungsqualität aufweisen (für Details, siehe Liepmann und Müller, 2018).

Dies betrifft zum einen Personen, die im Jahr 1989 älter als 60 waren. Hier ist es wahrscheinlich, dass viele dieser Personen zwischen 1989 und 1992 den Arbeitsmarkt verlassen haben und deshalb nicht mehr in den IEB erschienen sind.

Zum anderen weisen unverheiratet Frauen in den GAV-Daten eine verhältnismäßig schlechte Verknüpfungsqualität mit den IEB auf. Dies liegt mutmaßlich in erster Linie daran, dass es in dieser Gruppe zwischen 1989 und 1992 relativ häufig zu Wechseln des Nachnamens aufgrund von Eheschließungen gekommen ist. Bei Analysen der Erwerbsverläufe von Frauen sollte deshalb besonders auf den Ehestatus geachtet werden. Es könnte zum Beispiel angebracht sein, den Familienstand als Kontrollvariable zu berücksichtigen oder zu untersuchen, ob sich die Analyseergebnisse nach Familienstandkategorie unterscheiden.

Hinsichtlich des Analysepotentials der im GAV-ADIAB 7519 enthaltenen Merkmale sind insbesondere folgende drei Einschränkungen hervorzuheben: Erstens konnten bei der Harmonisierung der Berufe und Wirtschaftszweige einige Werte nicht umgeschlüsselt werden, wegen fehlender Labels oder da es keinen entsprechenden Wirtschaftszweig in den IEB gab.

Zweitens folgt die Klassifizierung der Wirtschaftszweige und Berufe im GAV einer anderen Logik als in den IEB, weshalb diese Kategorien oftmals nicht direkt vergleichbar sind. Trotz großer Sorgfalt bei der Umschlüsselung anhand der Wirtschaftszweig- bzw. Berufsbezeichnungen ist davon auszugehen, dass nicht alle umgeschlüsselten Kategorien vollständig vergleichbar sind. Hierauf ist bei Analysen auf Wirtschaftszweig- und Berufsebene zu achten, z. B. indem Berufe noch stärker aggregiert werden. Zusätzlich wird die Variable Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau) in den IEB erst mit der KldB2010 verfügbar und ist somit erst ab 2012 mit tatsächlichen Angaben gefüllt. Werte in früheren Jahren wurden imputiert und bei Analysen ist entsprechende Vorsicht geboten.

Drittens befindet sich der Gebietsstand für den harmonisierten Arbeits- und Wohnort in den GAV-Daten auf dem Stand vom 31.12.2014, da dieser zum Zeitpunkt der Umschlüsselung aktuell war, während die IEB den Gebietsstand vom 31.12.2019 aufweisen. Hierauf ist bei Analysen auf regionaler Ebene zu achten, z. B. indem von Gebietsstandänderungen betroffene Kreise zusammengefasst werden.

4.2 Gesamte IEB

4.2.1 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 6 erhebt keinen Anspruch auf Vollständigkeit.

Tabelle 6: Biographische Lücken und Identifikationsmöglichkeiten

Biografische Lücke	**Information über Lücke, ggf. identifizierbar anhand der Ausprägung des Merkmals „grund“ im Vorgängersatz der Quelle**
Beamte/Beamtinnen, Berufssoldaten/-soldatinnen, 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 aber unter bestimmten Bedingungen in der Quelle ASU abgebildet, vgl. Abschnitt 3.2.1.4)	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-)Landwirte/-wirtinnen	
Pflegepersonen i.S. von § 19 SBG XI	
Grundwehrdienstleistende	BeH, LeH, LHG, ASU, XASU
Wehrübungsleistende	BeH, LeH, LHG, ASU, XASU
Zivildienstleistende	BeH, LeH, LHG, ASU, XASU
Ableistende eines freiwilligen sozialen bzw. ökologischen Jahres anstelle des Zivildienstes	
Sonstige Personen, die nicht bei der Rentenversicherung oder der BA registriert sind (z.B. Sabbatjahr, Finanzierung aus privatem Vermögen oder Rentenbezügen, Emigration, Beschäftigung im Ausland, ehrenamtliche Tätigkeit, etc.)	BeH, LeH, ASU
Streikende bei Streikdauer über einem Monat	LeH
Sozialhilfe-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)	

4.2.2 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 Kapiteln zu den Datenquellen.

4.3 Beschäftigtenhistorik

- Die Übermittlung von Informationen zur Ausbildung, ausgeübten Tätigkeit und Stellung im Beruf erfolgt durch Meldungen der Arbeitgeber nach DEÜV (siehe Abschnitt 2.1) 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.[6]

- 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 Sozialversicherung eingeräumt wurde,[7] zu einer temporären Zunahme fehlender Angaben. Analysen der Statistik der BA (Bertat et al., 2013, S. 10) zeigen, dass die neuen bzw. umgestellten Merkmale „ausgeübte Tätigkeit“, „Arbeitszeit“ und „Ausbildungsabschluss“ nach der Umstellung zu etwa 20% bis 30% keine Angaben enthielten. Eine deutliche Verbesserung stellt sich im ersten Halbjahr 2013 ein. Um die Qualität der Arbeitszeitangabe im Übergangszeitraum zu verbessern, haben Ludsteck und Thomsen (2016) ein Imputationsverfahren entwickelt, das die fehlenden durch imputierte Werte ersetzt. Die imputierten Daten werden im GAV-ADIAB 7519 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 Jahrs 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 vom 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

[6] Siehe Paulus und Matthes, 2013, für Details zur Klassifikation der Berufe 2010.

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

Februar (ab Jahresmeldung 2013) des folgenden Jahres bestätigt werden. Faktisch treffen Meldungen jedoch zum Teil um Jahre verspätet ein. Die Historikdatei des IAB wird aber nicht kontinuierlich fortgeschrieben, sondern in bestimmten Abständen aktualisiert. Dies geschieht unter Verwendung von Dateien mit den Beschäftigten-Meldungen jeweils eines Jahres, die 36, 18, 12 oder 6 Monate nach Ablauf des Berichtsjahres eingetroffen sind (frühestens im Juli 2015 z.B. kann die 18-Monats-Datei für das Jahr 2013 erstellt werden). Mehr als 3 Jahre verspätete Meldungen werden im IAB nicht berücksichtigt, so dass eine 36-Monats-Datei per Definition einen Füllgrad von 100% aufweist.

- In der zugrundeliegenden Version der IEB ist 2016 das letzte Jahr mit einem Füllgrad der BeH-Meldungen von 100%.[8] Für die Jahre 2017 und 2018 wurden 18-Monats-Dateien verwendet, für 2019 eine 6-Monats-Datei. Es ist daher davon auszugehen, dass Beschäftigungsmeldungen für die Jahre 2017 bis 2018 leicht und für 2019 etwas stärker untererfasst sind. Die Auswertbarkeit der Daten auf Personenebene sollte dadurch aber nicht eingeschränkt sein. Allerdings häufen sich die fehlenden Meldungen in wenigen Betrieben. Dies bedeutet, dass in Einzelfällen die Betriebsangaben, z.B. die Betriebsgröße, grob falsch sind und sich in Folgeversionen stark verändern werden.
- Im Jahr 1984 erfolgte eine Änderung des Meldeverfahrens. Einmalig gezahltes Bruttoarbeitsentgelt wurde ab diesem Zeitpunkt als Teil des beitragspflichtigen Jahresentgelts gemeldet, was zu einer Erhöhung des durchschnittlichen Tagesentgelts führt. Insbesondere steigt ab diesem Jahr der Anteil der Entgelte oberhalb der Beitragsbemessungsgrenze stark an (vgl. Bender et al., 1996).
- Für die Jahre 1992 bis 2000 wurden in den IEB auffällige Einbrüche und Anstiege in der Anzahl der Meldungen festgestellt. Es lassen sich v.a. Einbrüche bei folgenden zehn Kreisen beobachten: Braunschweig (03101), Wolfsburg (03103), Emden (03402), Kassel (06633), Essen (05113), Neuss (05162), Erftkreis (05362), Hersfeld-Rotenburg (06632), Miltenberg (09676) und Kempten (Allgäu) (09763). Grund hierfür sind Meldeprobleme bei einzelnen bzw. mehreren Betrieben in diesen Regionen.
- Nennenswerte Einbrüche verzeichnen auch die Kreise Salzgitter (03102) und Hoyerswerda (14264).
- Bei den Meldungen zu Vollzeitbeschäftigung in den IEB sind besonders die Kreise Main-Taunus-Kreis (06436) und Alzey-Worms (07331) erwähnenswert. Hier sind überdurchschnittliche Anstiege zu beobachten. Als Grund sind ebenfalls Meldeprobleme bei einzelnen bzw. mehreren Betrieben in diesen Regionen anzuführen.
- In den Jahren 1996 bis 1998 kommen die Werte 841-844 (Ärzte und Apotheker) in der Variable Beruf – ausgeübte Tätigkeit (beruf) im Vergleich zu benachbarten Jahren sehr selten vor. Die Gründe dafür sind nicht bekannt.
- 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

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

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.

4.4 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 erhalten. 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.

4.5 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 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% der Fälle. In XSozial kommt es zu trägerspezifischen Untererfassungen dieser Personengruppe. Im GAV-ADIAB 7519 ist der Ausschluss von Leistungen nicht identifizierbar.
- Aufgrund der Meldelogik sind Informationen aus dem Meldeverfahren XSozial nur monatsgenau zuspielbar.
- 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.

4.6 Arbeitsuchendenhistoriken (ASU/XASU)

4.6.1 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 arbeitsuchende Personen nicht mehr vollständig mit BA-Verfahren erfasst. Die Quelle ASU berücksichtigt ab diesem Zeitpunkt nur noch Personen, die von der BA im Rechtskreis SGB III (Arbeitsförderung), oder von ARGEn, gE bzw. gT im Rechtskreis SGB II (Grundsicherung) betreut werden.
- Für die Vermittlungskräfte ist es nicht möglich, die Zuordnung zum Rechtskreis immer unmittelbar zu erfassen, da vielfach erst zeitverzögert feststeht, welcher Träger aufgrund einer möglichen SGB-II-Leistungsberechtigung vorrangig zuständig ist. Es wird daher empfohlen, die Ausprägung des Merkmals „Trägerart“ in der ASU mit der Ausprägung in der LHG im gleichen Zeitraum abzugleichen. Aufgrund der Erfassungslücken bei LHG im Zeitraum 2005 bis 2006 ist dies nicht immer möglich.
- Für manche Personen, für die ab 2005 eine Optionskommune zuständig war, wurden von der BA parallel "künstliche" ASU-Sätze angelegt. Diese sind über estatvor (Wechsel zu einem zugelassenen kommunalen Träger) zu identifizieren.
- Von Mitte 2005 bis Mitte 2006 wurde in der BA das EDV-Verfahren coArb, aus dem die Arbeitsuchendendaten 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 Arbeitsuchenden. Lückenhaft erfasst wurden auch Daten zu Personen, die sich lediglich beraten lassen wollten. Die Daten der Berufsberatung wurden in einem separaten System gehalten. Mit VerBIS erweiterten sich die Ausprägungen des Arbeitsuche-Status um „Ratsuchende“ und Personen „ohne Status“. Zu Letzteren gehören ALG-II-Leistungsberechtigte, die 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.

4.6.2 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, arbeitsuchend“ 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.

4.7 Maßnahmeteilnahmehistoriken (MTH/XMTH)

4.7.1 MTH

- Die MTH sind für Maßnahmen mit einem Eintrittsdatum vor dem 01.01.2000 unvollständig.
- Zum 01.01.2005 kommt es zu einem Bruch, da mit Einführung des SGB II Maßnahmeteilnehmende auf verschiedene Träger aufgeteilt wurden (siehe Abschnitte 2.2.5 und 4.2.2).
- 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 4.2.2) kann es zur Aufspaltung der Dokumentation von Maßnahmeteilnahmen in der MTH und XMTH kommen, wenn es zu einer Umstellung des Meldeverfahrens kommt.[9] Dies kann sich in einem Split des Maßnahmespells, aber auch in Dopplungen niederschlagen (siehe Abschnitt 3.2.1.6).
- 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.

4.7.2 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

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

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 4.2.2) 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 3.2.1.6).
- Die über monatliche Stichtagsauszählungen in der XMTH erhaltenen Zahlen bzw. Personen unterscheiden sich in mehrfacher Hinsicht von den veröffentlichten BA-Statistiken.

5 Variablenbeschreibung

Auszählungen sowie Übersichten über die einzelnen Ausprägungen und Labels der Merkmale finden sich in den Arbeitshilfen zum GAV-ADIAB 7519 unter https://fdz.iab.de.

5.1 Identifikatoren

5.1.1 Systemfreie Personennummer (persnr)

Kategorie	**Beschreibung**
Variablenlabel	Systemfreie Personennummer
Variablenname	persnr
Kategorie	Identifikatoren
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH, GAV
Datentyp	Numerisch
Detailbeschreibung	Die systemfreie Personennummer zeigt an, welche Sätze zur gleichen Person gehören. Systemfrei bedeutet, dass von dieser Personennummer nicht auf Eigenschaften der Person oder Original-Identifikatoren geschlossen werden kann. Da es in den verschiedenen Datenquellen keinen einheitlichen Personen-Identifikator gibt, ist die Zuordnung der Informationen aus den unterschiedlichen Datenquellen (z.B. Beschäftigung und Leistung) zu Personen(-Konten) nicht immer eindeutig. In solchen Fällen kann es zu unplausiblen Erwerbshistorien kommen. Die Bildung des quellenübergreifenden Personen-Identifikators basiert auf einer von der BA entwickelten Heuristik.

5.1.2 Systemfreie Betriebsnummer GAV-ADIAB (betnr_gavadiab)

Kategorie	**Beschreibung**
Variablenlabel	Systemfr. Betriebnr. GAV-ADIAB
Variablenname	betnr_gavadiab
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/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. 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

Kategorie	Beschreibung
	Niederlassung ein Betrieb und erhält je eine Betriebsnummer, wenn sie **nicht** derselben Wirtschaftsklasse angehören. Hat das Unternehmen **mehrere** Niederlassungen in **mehreren** Gemeinden, so ist jede dieser Niederlassungen ein Betrieb und erhält je eine Betriebsnummer. In diesem Zusammenhang sind auch folgende Definitionen im Sinne der Betriebsnummernvergabe im Rahmen des Meldeverfahrens zur Sozialversicherung zu beachten: 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. Eine **Arbeitsstätte** ist eine Einheit, in der Beschäftigte tätig sind und für die nach o.a. Grundsätzen keine Betriebsnummer zu vergeben ist. Ein **Unternehmen** als Begriff fasst Betriebs- und Arbeitsstätten desselben Arbeitgebers zusammen. Ein **Arbeitgeber** ist jede natürliche oder juristische Person, für die mindestens ein sozialversicherungspflichtig bzw. geringfügig Beschäftigter tätig ist. Betrieb und Betriebsstätten sind synonyme Begriffe; Niederlassung ist dann ein synonymer Begriff für Filialen, Zweigstellen, Auslagerungen, Dependancen, Arbeitsstätten u.ä., wenn sie kein Betrieb ist.
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.

5.2 Generierte technische Merkmale

5.2.1 Teil der GAV Stichprobe (gav)

Kategorie	Beschreibung
Variablenlabel	Teil der GAV Stichprobe
Variablenname	gav
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH, GAV
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal gibt Auskunft darüber, ob die Sätze zu einer Person gehören, die Teil der GAV-Stichprobe oder der Vergleichsgruppe ist (0 - Nicht in GAV-Stichprobe (1989 BRD); 1 - In GAV-Stichprobe (1989 DDR)).

5.2.2 Satzzähler pro Konto (spell)

Kategorie	Beschreibung
Variablenlabel	Satzzähler pro Konto
Variablenname	spell
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch

Kategorie	Beschreibung
Detailbeschreibung	Der Satzzähler pro Konto nummeriert aufsteigend die IEB-Sätze einer Person. Das Merkmal wird beim Episodensplitting erzeugt und bezieht sich auf die gesplitteten Sätze. Mit dem Merkmal „Satzzähler pro Konto“ kann die ursprüngliche Sortierung auf einfache Art und Weise wiederhergestellt werden. Die Sortierung erfolgt grundsätzlich zunächst nach dem Beginndatum der Episode und dann nach der Quelle. Innerhalb der Beschäftigungsmeldungen werden sozialversicherungspflichtige vor geringfügige Beschäftigungsmeldungen und höhere vor niedrigere Tagesentgelte sortiert. Eine Ausnahme bilden Einmalzahlungen, die nach hinten sortiert werden.

5.2.3 Quelle des Satzes (quelle)

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

5.2.4 Jahr (jahr)

Kategorie	Beschreibung
Variablenlabel	Jahr
Variablenname	jahr
Kategorie	Generierte technische Merkmale
Herkunft	BHP
Datentyp	Numerisch

Detailbeschreibung	Dieses Merkmal ist nur in der Basisbetriebsdatei enthalten. Es gibt das Gültigkeitsjahr der Betriebsinformationen zum Stichtag 30.06 an.
	Mit Hilfe dieses Merkmals und der Betriebsnummer können Personen- und Betriebsdatei verknüpft werden.
	Beispielcode mit dem „merge“-Befehl in Stata 16:
	`use "${orig}/GAV-ADIAB_7519_v1"`
	`gen jahr = year(begepi)`
	`label variable jahr Jahr`
	`sort betnr_gavadiab jahr`
	`merge m:1 betnr_gavadiab jahr ///`
	`using "${orig}/GAV-ADIAB_7519_v1_bhp_basis_v2"`

5.3 Gültigkeitszeitraum

5.3.1 Beginndatum Originalbeobachtung (begorig)

Kategorie	Beschreibung
Variablenlabel	Beginndatum Originalbeobachtung
Variablenname	begorig

Kategorie	Beschreibung
Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Datum
Detailbeschreibung	Das Beginndatum des Originalsatzes entspricht dem ursprünglichen Beginndatum der Meldung. Dies kann vom Beginndatum der betreffenden Datenzeile, dem sogenannten Episoden-Beginndatum, abweichen (vgl. auch die Erläuterungen zum Episodensplitting in Abschnitt 3.2.3). Da die Meldelogik unter bestimmten Umständen eine Re-identifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben in diesen Fällen durch das in Abschnitt 3.2.2 beschriebene Anonymisierungsverfahren verändert. **1) BeH** 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. **2) LHG, ASU, XASU** Bestimmte Änderungen führen zur Anlage eines neuen Datensatzes (siehe Abschnitte 3.2.1.3 und 3.2.1.4). begorig gibt das Beginndatum des neuen Zeitraums an. **3) MTH, XMTH** In der MTH und XMTH entstehen neue Datensätze, wenn ein Trägerwechsel während der Teilnahme an einer arbeitsmarkpolitischen Maßnahme stattfindet.

5.3.2 Enddatum Originalbeobachtung (endorig)

Kategorie	Beschreibung
Variablenlabel	Enddatum Originalbeobachtung
Variablenname	endorig
Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Datum
Detailbeschreibung	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 3.2.3). Da die Meldelogik unter bestimmten Umständen eine Re-identifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben in diesen Fällen durch das in Abschnitt 3.2.2 beschriebene Anonymisierungsverfahren verändert. **1) BeH** 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. **2) LHG, ASU, XASU** Bestimmte Änderungen führen zur Anlage eines neuen Datensatzes (siehe Abschnitte 3.2.1.3 und 3.2.1.4). endorig gibt das Enddatum des neuen Zeitraums an. **3) MTH, XMTH**

Kategorie	Beschreibung
	In der MTH und XMTH entstehen neue Datensätze, wenn ein Trägerwechsel während der Teilnahme an einer arbeitsmarktpolitischen Maßnahme stattfindet.

5.3.3 Beginndatum Episode (begepi)

Kategorie	Beschreibung
Variablenlabel	Beginndatum Episode
Variablenname	begepi
Kategorie	Generierter Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Datum
Detailbeschreibung	Das Beginndatum der Episode ist immer gleich dem oder größer als das Beginndatum der Originalbeobachtung (vgl. auch die Erläuterungen zum Episodensplitting in Abschnitt 3.2.3).

5.3.4 Enddatum Episode (endepi)

Kategorie	Beschreibung
Variablenlabel	Enddatum Episode
Variablenname	endepi
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 3.2.3).

5.4 Informationen zur Person

5.4.1 Geschlecht (frau)

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

5.4.2 Geburtsjahr (gebjahr)

Kategorie	Beschreibung
Variablenlabel	Geburtsjahr
Variablenname	gebjahr
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH, GAV

Kategorie	Beschreibung
Datentyp	Numerisch
Detailbeschreibung	Das Geburtsjahr ist innerhalb eines Personenkontos konstant.
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.

5.4.3 Geburtsmonat (gebmon)

Kategorie	Beschreibung
Variablenlabel	Geburtsmonat
Variablenname	gebmon
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH, GAV
Datentyp	Numerisch

Detailbeschreibung	Der Geburtsmonat ist innerhalb eines Personenkontos konstant.
	Um aus den Angaben zum Geburtsjahr (gebjahr) und -monat (gebmon) eine Variable im Datumsformat JJJJmM (z.B. 1984m6) zu generieren, kann folgender Befehl verwendet werden:
	`gen int gebdat = ym(gebjahr, gebmon)`
	`format gebdat %tm`

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.

5.4.4 Staatsangehoerigkeit, vergroebert (deutsch)

Kategorie	Beschreibung
Variablenlabel	Staatsangehoerigkeit, vergroebert
Variablenname	deutsch
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH, GAV
Datentyp	Numerisch
Detailbeschreibung	**1) BeH, LeH, LHG, ASU, XASU, MTH, XMTH** Das Merkmal gibt an, ob die entsprechende Person die deutsche Staatsbürgerschaft (1) hat oder nicht (0). Diese Angabe kann sich im Zeitverlauf ändern. **2) GAV** Das Merkmal gibt an, ob die entsprechende Person die deutsche Staatsbürgerschaft hat (1) oder nicht (0). Die Staatsangehörigkeiten wurden den Staatsangehörigkeiten der BRD basierend auf dem Namen des Staates zugewiesen.
Qualitätshinweis	Das Merkmal ist in der LeH erst ab 1983 gut gefüllt.
Besonderheiten	Eine weniger stark vergröberte Bereitstellung dieser Variable ist aufgrund der geringen Anzahl an Personen ohne deutsche Staatsangehörigkeit in der GAV-Stichprobe nicht möglich.

5.4.5 Familienstand (famst)

Kategorie	Beschreibung
Variablenlabel	Familienstand
Variablenname	famst
Kategorie	Persönliche Merkmale
Herkunft	LeH, LHG, ASU, XASU, MTH, XMTH, GAV
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal beschreibt den Familienstand. Diese Variable ist für Beschäftigungsmeldungen nicht gefüllt! **1) LeH** In der Quelle LeH umfasst das Merkmal nur zwei Ausprägungen (0 - nicht verheiratet, 1 - verheiratet). **2) LHG, ASU, XASU, MTH, XMTH** In den Quellen LHG/ASU/XASU/MTH/XMTH wird zwischen sechs Ausprägungen (Werte 11-16) unterschieden. **3) GAV** In der Quelle GAV wird zwischen vier Ausprägungen (Werte 111-114) unterschieden. Ein Abgleich der Informationen aus den Quellen wurde nicht vorgenommen.
Qualitätshinweis	**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%. **2) LHG, XASU, XMTH** Aufgrund abweichender Meldestandards in XSozial sind die Angaben aus den zkT bis Dezember 2009 als weniger valide einzustufen.

5.4.6 Kinderzahl (kind)

Kategorie	Beschreibung
Variablenlabel	Kinderzahl
Variablenname	kind
Kategorie	Persönliche Merkmale
Herkunft	LeH, LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	Je nach Datenquelle hat das Merkmal eine unterschiedliche Bedeutung. Diese Variable ist für Beschäftigungsmeldungen nicht gefüllt! **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 **2) ASU, MTH**

Kategorie	Beschreibung
	Hier entspricht die Ausprägung des Merkmals der Anzahl an Kindern unter 15 Jahren im Haushalt. **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.
Qualitätshinweis	**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. **3) 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.

5.4.7 Ausbildung (ausbildung)

Kategorie	Beschreibung
Variablenlabel	Ausbildung
Variablenname	ausbildung
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält den beruflichen Ausbildungsabschluss. Eine Einordnung der Abschlüsse in die Internationale Standardklassifikation des Bildungswesens (ISCED 2011) ist den zentralen Arbeitshilfen des FDZ zu entnehmen. Zu beachten ist, dass das Merkmal je nach Quelle eine unterschiedliche Bedeutung hat. **1) BeH** Das Merkmal enthält für BeH-Sätze die von den Arbeitgebern im Rahmen des Meldeverfahrens anzugebende Berufsausbildung. Folgende Ausprägungen kommen vor: 1 ohne abgeschlossene Berufsausbildung 2 betriebliche/außerbetriebliche Ausbildung 11 FH ohne nähere Spezifikation 12 Hochschulabschluss ohne nähere Spezifikation Bei Meldungen nach dem neuen Tätigkeitsschlüssel (siehe Abschnitt 4.3) ist es nicht mehr möglich, FH-Absolventinnen und -Absolventen wie bisher eindeutig zu identifizieren, da der neue Tätigkeitsschlüssel die separate Erfassung dieses Bildungsabschlusses nicht länger erlaubt. Sie werden der Kategorie 12 zugeordnet. **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: 1 ohne abgeschlossene Berufsausbildung 2 betriebliche/außerbetriebliche Ausbildung 3 Berufsfachschule

Kategorie	Beschreibung
	4 Fachschule 5 Fachhochschule 6 Universität 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 Aufprägungen vorliegen: 7 in Deutschland nicht anerkannte Berufsausbildung 8 in Deutschland nicht anerkannter Hochschulabschluss 9 betriebliche/schulische Ausbildung 10 sonstige Examen 11 FH ohne nähere Spezifikation 12 Hochschulabschluss ohne nähere Spezifikation 13 Promotion 14 Bachelor (BA) 15 Bachelor (FH) 16 Bachelor (Uni) 17 Master (FH) 18 Master (Uni) 19 Diplom (BA) 20 Diplom (FH) 21 Diplom (Uni) 23 Grundständige Studienfächer/-gänge 24 Weiterführende Studienfächer/-gänge 25 Sonstige (Aufstiegs-)Weiterbildungen für Hochschulabsolventen **3) XASU, XMTH** Für diese Sätze ist die letzte abgeschlossene Berufsausbildung enthalten. Folgende Ausprägungen kommen vor: 1 ohne abgeschlossene Berufsausbildung 2 betriebliche/außerbetriebliche Ausbildung 3 Berufsfachschule 4 Fachschule 5 Fachhochschule 6 Universität 7 in Deutschland nicht anerkannte Berufsausbildung 8 in Deutschland nicht anerkannter Hochschulabschluss Für Auswertungen über einen längeren Zeitraum können die Ausprägungen folgendermaßen aggregiert werden: 1, 22 auf 1 7 auf 2 8 auf 3 2, 3, 4, 9 auf 4 5, 11, 14, 15, 17, 19, 20 auf 5

Kategorie	Beschreibung
	6, 10, 12, 13, 16, 18, 21 auf 6 23, 24, 25 auf 7 Die aggregierten Werte haben folgende Labels: 1 ohne (anerkannte) abgeschlossene Berufsausbildung 2 in Deutschland nicht anerkannte Berufsausbildung 3 in Deutschland nicht anerkannter Hochschulabschluss 4 betriebliche / schulische Ausbildung 5 Fachhochschule 6 Universität 6 sonstige Examen 7 Grundständige Studienfächer/-gänge 7 Weiterführende Studienfächer/-gänge 7 Sonstige (Aufstiegs-)Weiterbildungen für Hochschulabsolventen
Qualitätshinweis	**1) BeH** „Zusammen mit einem Betriebswechsel treten häufig auch Änderungen im Ausbildungsstatus auf. Das liegt daran, dass beim neuen Betrieb die Meldedaten neu zusammenstellt werden. Wenn eine Beschäftigte oder 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 und Koch, 2004, S. 63). Der Anteil der fehlenden Werte steigt in der BeH im Zeitverlauf fast kontinuierlich an. Durch die Einführung des neuen Tätigkeitsschlüssels gab es im Jahr 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 folgenden Personengruppen auf: geringfügig Beschäftigte, Personen in Teilzeit, ausländische Beschäftigte und ostdeutsche Arbeitnehmerinnen und Arbeitnehmer. Hintergrund ist, dass das Merkmal versicherungsrechtlich keine besondere Bedeutung hat (siehe Meinken und Koch, 2004, S. 63). 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. Für das Merkmal Ausbildung imputiert (ausbildung_imp, siehe Abschnitt 5.4.8) 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. **2) ASU, MTH** 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 Aufprägungen zu finden. **2) XASU, XMTH** Der Füllgrad in den Quellen XASU und XMTH ist generell niedrig.

5.4.8 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 5.4.7 beschriebenen Probleme der Variable Ausbildung (ausbildung) bei BeH-Meldungen. Das Imputationsverfahren wird in Thomsen et al. (2018) beschrieben, welches auf der Arbeit von Fitzenberger et al. (2006) basiert. 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). Eine Einordnung der Abschlüsse in die Internationale Standardklassifikation des Bildungswesens (ISCED 2011) ist den zentralen Arbeitshilfen des FDZ zu entnehmen.
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.

5.4.9 Schulausbildung (schule)

Kategorie	Beschreibung
Variablenlabel	Schulausbildung
Variablenname	schule
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält den Schulabschluss. Eine Einordnung der Abschlüsse in die Internationale Standardklassifikation des Bildungswesens (ISCED 2011) ist den zentralen Arbeitshilfen des FDZ zu entnehmen. Je nach Quelle sind unterschiedliche Ausprägungen möglich. **1) BeH** Mit der Umstellung des Tätigkeitsschlüssels im Jahr 2011 ändern sich die möglichen Ausprägungen (siehe Abschnitt 4.3). Davor gelten: 5 Volks-/Hauptschule o./m. Abschluss, Mittlere Reife oder gleichwertige Schulbildung 8 Fachhochschulreife oder Abitur/Hochschulreife 9 Abitur/Hochschulreife Danach gelten:

Kategorie	Beschreibung
	1 Kein Hauptschulabschluss 4 Hauptschulabschluss/Volksschulabschluss 6 Mittlere Reife 8 Fachhochschulreife oder Abitur/Hochschulreife 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). **2) ASU, XASU, MTH, XMTH** Für diese Quellen sind die folgenden Ausprägungen möglich: 1 Kein Hauptschulabschluss 4 Hauptschulabschluss/Volksschulabschluss 6 Mittlere Reife 7 Fachhochschulreife 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	**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. **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.

5.5 Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

5.5.1 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	**1) BeH** In BeH-Sätzen weist das Merkmal jeweils das Bruttotagesentgelt einer oder eines Beschäftigten aus. Es wird berechnet aus dem vom Arbeitgeber gemeldeten Zeitraumentgelt sowie der Dauer des Zeitraums der (ungesplitteten) Originalmeldung in Kalendertagen. Das Tagesentgelt ist in Euro angegeben. Bis 1998 meldeten Arbeitgeber grundsätzlich nur das sozialversicherungspflichtige Arbeitsentgelt. Entgelte, die unterhalb der Geringfügigkeitsgrenze lagen, wurden

Kategorie	Beschreibung
	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. 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.) Die Geringfügigkeits- und Beitragsbemessungsgrenzen unterscheiden sich von Jahr zu Jahr und in West- und Ostdeutschland (maßgeblich ist jeweils der Standort des Betriebs). Eine Übersicht über diese Grenzen findet sich unter https://fdz.iab.de. 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. Ab 2013 nimmt die Anzahl der Meldungen mit Abmeldegrund 54 (Meldung eines einmalig gezahlten Entgelts) sprunghaft zu (siehe Abschnitt 5.5.12). Es ist davon auszugehen, dass Sonderzahlungen, die vor 2013 mit den Jahresmeldungen angegeben wurden, nun extra gemeldet werden. Daher empfiehlt es sich bei Analysen der Entgelte im Zeitverlauf, die Sonderzahlungen zu den entsprechenden Entgelten von zeitgleichen Beschäftigungsepisoden im gleichen Betrieb dazu zu addieren. 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. **2) LeH** 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. Die Angabe von 0 Euro ist seit 01.01.2005 auf Sperrzeiten oder Leistungsunterbrechungen zurückzuführen. Ist für einen Satz mit Leistungssatz gleich 0 ein Beendigungsgrund angegeben, so handelt es sich um eine Unterbrechungsmeldung. Bei Sätzen, die eine Sperrzeit widerspiegeln, ist der Anspruch derselbe wie vor Beginn der Sperrzeit.

5.5.2 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, GAV
Datentyp	Numerisch
Detailbeschreibung	**1) BeH** 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 eine 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). Der Arbeitgeber verschlüsselt die Tätigkeit der oder des Beschäftigten hierbei gemäß der „Klassifizierung der Berufe 1988“ (Bundesanstalt für Arbeit, 1988), in der ca. 25.000 Berufsbezeichnungen zu finden sind. Die Berufsordnung besteht aus einer 3-stelligen Kennziffer und umfasst rund 330 Ausprägungen. Beschäftigungsmeldungen mit einem Enddatum nach dem 30.11.2011 werden nach dem neuen Tätigkeitsschlüssel 2010 (KldB2010) gemeldet (siehe Abschnitt 4.3). Diese Werte werden durch einen Schwerpunktumstieg auf die KldB1988 zurückgeschlüsselt. Dabei kommt es zu Unschärfen. **2) ASU, XASU, MTH** Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1). **3) GAV** Dieses Merkmal basiert auf den Volkswirtschaftlichen Arbeitskräftesystematiken. Zur Vergleichbarkeit mit der IEB Berufsklassifikation wurde für jeden 8-stelligen Berufscode anhand des Vergleiches der Wertelabels ein äquivalenter Berufscode in der KldB 2010 zugewiesen. Dadurch wurden 3.359 GAV Berufscodes 1.268 KldB 2010 Berufscodes zugewiesen. Anschließend wurden diese 1.268 KldB 2010 Berufscodes mit einer Umschlüsselungstabelle der Bundesagentur für Arbeit auf 334 KldB 1988 Berufscodes umgeschlüsselt.
Qualitätshinweis	**1) BeH** Im Jahr 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. **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%. **3) XASU** Das Merkmal Beruf ist nahezu für den gesamten zur Verfügung stehenden Zeitraum nicht gefüllt. **4) GAV** Für beinahe 15% der Beobachtungen im GAV weist das Merkmal fehlende Werte auf. Fast 75% dieser fehlenden Werte sind bedingt durch fehlende Werte bereits im GAV Berufscode.

5.5.3 Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (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, GAV
Datentyp	Numerisch
Detailbeschreibung	**1) BeH** Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V). Der Arbeitgeber verschlüsselt die Tätigkeit der/des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Die weniger detaillierte Berufsgruppe wird durch die ersten drei Stellen der Kennziffer erfasst. Das in der KldB2010 an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt. Beschäftigungsmeldungen mit einem Enddatum vor dem 30.11.2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet (siehe Abschnitt 4.3). Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten. **2) LeH, ASU, XASU, MTH** Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1). **3) GAV** Dieses Merkmal basiert auf den Volkswirtschaftlichen Arbeitskräftesystematiken. Zur Vergleichbarkeit mit der IEB Berufsklassifikation wurde für jeden 8-stelligen Berufscode anhand des Vergleiches der Wertelabels ein äquivalenter Berufscode in der KldB 2010 zugewiesen. Dadurch wurden 3.359 GAV Berufscodes 1.268 KldB 2010 Berufscodes zugewiesen. Anschließend wird die 5-stellige KldB 2010 auf den 3-Steller gekürzt.
Qualitätshinweis	**1) BeH** Im Jahr 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. **2) LeH** Das Merkmal wurde früher für die Quelle LeH nicht gefüllt, da die Berufsangaben in der LeH eine schlechte Qualität aufwiesen. Mittlerweile wird der Beruf in der LeH aus der Arbeitsuchendenhistorik zugespielt, so dass die Qualität gestiegen ist. **3) XASU** Das Merkmal ist für den gesamten zur Verfügung stehenden Zeitraum sehr schlecht gefüllt.

Kategorie	Beschreibung
	4) GAV Für beinahe 15% der Beobachtungen im GAV weist das Merkmal fehlende Werte auf. Fast 75% dieser fehlenden Werte sind bedingt durch fehlende Werte bereits im GAV Berufscode.

5.5.4 Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (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
Datentyp	Numerisch
Detailbeschreibung	**1) BeH** 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 eine 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). Der Arbeitgeber verschlüsselt die Tätigkeit der oder des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Die weniger detaillierte Berufsuntergruppe wird durch die ersten vier Stellen der Kennziffer erfasst. Das in der KldB2010 an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt. Beschäftigungsmeldungen mit einem Enddatum vor dem 30.11.2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet (siehe Abschnitt 4.3). Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten. **2) LeH, ASU, XASU, MTH** Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1).
Qualitätshinweis	**1) BeH** Im Jahr 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. **2) LeH** Das Merkmal wurde früher für die Quelle LeH nicht gefüllt, da die Berufsangaben in der LeH eine schlechte Qualität aufwiesen. Mittlerweile wird der Beruf in der LeH aus der Arbeitsuchendenhistorik zugespielt, so dass die Qualität gestiegen ist. **3) XASU** Das Merkmal ist für den gesamten zur Verfügung stehenden Zeitraum sehr schlecht gefüllt.

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

5.5.5 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, GAV
Datentyp	Numerisch
Detailbeschreibung	**1) BeH** 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 eine 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). Der Arbeitgeber verschlüsselt die Tätigkeit der oder des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Das an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt. Beschäftigungsmeldungen mit einem Enddatum vor dem 30.11.2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet (siehe Abschnitt 4.3). Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten. **2) LeH, ASU, XASU, MTH** Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel siehe 1). **3) GAV** Dieses Merkmal basiert auf den Volkswirtschaftlichen Arbeitskräftesystematiken. Zur Vergleichbarkeit mit der IEB Berufsklassifikation wurde für jeden 8-stelligen Berufscode anhand des Vergleiches der Wertelabels ein äquivalenter Berufscode in der KldB 2010 zugewiesen. Dadurch wurden 3.359 GAV Berufscodes 1.268 KldB 2010 Berufscodes zugewiesen. Das an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt.
Qualitätshinweis	**1) BeH** Im Jahr 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. **2) LeH**

Kategorie	Beschreibung
	Das Merkmal wurde früher für die Quelle LeH nicht gefüllt, da die Berufsangaben in der LeH eine schlechte Qualität aufwiesen. Mittlerweile wird der Beruf in der LeH aus der Arbeitsuchendenhistorik zugespielt, so dass die Qualität gestiegen ist. **3) XASU** Das Merkmal ist für den gesamten zur Verfügung stehenden Zeitraum sehr schlecht gefüllt. **4) GAV** Für beinahe 15% der Beobachtungen im GAV weist das Merkmal fehlende Werte auf. Fast 75% dieser fehlenden Werte sind bedingt durch fehlende Werte bereits im GAV Berufscode. Außerdem wird die Variable erst mit der KldB2010 verfügbar und ist somit erst ab 2012 mit tatsächlichen Angaben gefüllt. Werte in früheren Jahren wurden imputiert und bei Analysen ist entsprechende Vorsicht geboten.

5.5.6 Teilzeit (teilzeit)

Kategorie	Beschreibung
Variablenlabel	Teilzeit
Variablenname	teilzeit
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, GAV
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal Teilzeit (teilzeit) unterscheidet zwischen Vollzeit- und Teilzeitbeschäftigten. **1) BeH** Für die Quelle BeH ist dabei das Verhältnis zwischen der arbeitsvertraglich vereinbarten und der betriebsüblichen Arbeitszeit maßgeblich. **2) GAV** Für die Quelle GAV basiert das Merkmal auf der Kategorie der gesetzlichen Arbeitszeit.
Qualitätshinweis	**1) BeH** Im Jahr 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. 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 findet sich in Fitzenberger und Seidlitz (2020). **2) GAV** In seltenen Fällen haben Teilzeitbeschäftigte eine Arbeitszeit von 40 Stunden und mehr und Vollzeitbeschäftigte von weniger als 40 Stunden (siehe Beschreibung des Merkmals GAV_az). Dieses Merkmal ist für Lehrlinge (siehe Beschreibung des Merkmals GAV_stib) nicht gefüllt.

5.5.7 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	Die berufliche Stellung einer oder 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. Nur für Vollzeitbeschäftigte liefert die Variable tatsächlich Angaben zur beruflichen Stellung, wobei unter anderem zwischen vollzeitbeschäftigten Arbeitern und Angestellten sowie Auszubildenden unterschieden wird. Die Unterscheidung zwischen vollzeitbeschäftigten Arbeitern und Angestellten richtet sich ausschließlich nach der Art der Rentenversicherungspflicht (BfA für Angestellte und LVA für Arbeiter). Meister und Poliere werden nur dann in einer eigenen Kategorie erfasst, wenn sie in der Arbeiterrentenversicherung pflichtversichert sind. Die Zuordnung eines Meisters oder Poliers zu den Arbeitern oder Angestellten kann nur über den jeweiligen Rententräger erfolgen. Neben den Auszubildenden, Praktikanten und Volontären zählen auch Anlernlinge, Schülerinnen oder Schüler an Schulen des Gesundheitswesens und Teilnehmende an geförderten Maßnahmen zur beruflichen Fortbildung, Umschulung und betrieblichen Einarbeitung zur Kategorie Beschäftigte in beruflicher Ausbildung. 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). Mit der Einführung des neuen Tätigkeitsschlüssels (siehe Abschnitt 4.3) fiel diese Unterscheidung jedoch weg. Die Variable „stib“ ist daher nur für Meldungen gefüllt, die noch vor Einführung des neuen Tätigkeitsschlüssels abgegeben wurden.
Qualitätshinweis	Im Jahr 1991 kommt es aufgrund der Wiedervereinigung zu einer deutlichen Häufung der fehlenden Angaben.

5.5.8 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

Kategorie	Beschreibung
Detailbeschreibung	Das Merkmal nimmt für jede Datenquelle unterschiedliche Werte mit unterschiedlichen Bedeutungen an. **1) BeH** 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. 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. Im Meldeverfahren ist geregelt, dass Wechsel im Erwerbsstatus - z.B. die Übernahme einer oder eines Auszubildenden nach dem Ende der Berufsausbildung im Ausbildungsbetrieb - durch eine erneute Meldung erfolgen müssen. 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. 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. **2) LeH** Für LeH-Sätze ist im Merkmal Erwerbsstatus die gruppierte Leistungsart angegeben. Es kann also unterschieden werden, ob eine Person Arbeitslosengeld, Arbeitslosenhilfe oder Unterhaltsgeld bezieht oder ob Beiträge zur privaten Pflegeversicherung von der BA übernommen werden. Mit Einführung des Arbeitslosengeld II 2005 entfallen die Leistungen Arbeitslosenhilfe, Unterhaltsgeld und Beiträge zur privaten Pflegeversicherung. **3) LHG** Für LHG-Sätze zeigt das Merkmal Erwerbsstatus an, ob die Person minderjährig erwerbsfähig, volljährig erwerbsfähig oder nicht erwerbsfähig ab Altersrentengrenze geführt wird. Da die Meldelogik in vielen Fällen eine Re-identifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben durch das in Abschnitt 3.2.2 beschriebene Anonymisierungsverfahren verändert. **4) ASU, XASU** 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 Bewerbende, 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.

Kategorie	Beschreibung
	“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. Da die Meldelogik in vielen Fällen eine Re-identifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben durch das in Abschnitt 3.2.2 beschriebene Anonymisierungsverfahren verändert. **5) MTH, XMTH** 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.
Qualitätshinweis	**1) LHG** 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. 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 diese zu unterschiedlichen Zeitpunkten dokumentieren. Ein weiterer Grund können unterschiedliche Geburtstagsangaben für die zeitgleichen BG-Kundinnen und -Kunden sein, die sich auf den Erwerbsfähigkeitsstatus auswirken können. **2) ASU, XASU** Die Kategorien „ratsuchend“ und „nicht arbeitssuchend“ sind erst seit der Einführung von Verbis (siehe Abschnitt 4.6) 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.

5.5.9 Gleitzone (gleitz)

Kategorie	Beschreibung
Variablenlabel	Gleitzone
Variablenname	gleitz
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal ist erst ab 2003 und nur für BeH-Datensätze gefüllt. Es zeigt an, ob es sich bei der Beschäftigungsmeldung um eine Beschäftigung im Niedriglohnbereich innerhalb der sogenannten „Gleitzone“ bzw. ab Juli 2019 in einem "Übergangsbereich" handelt (sog. Midi-Job), für die die oder der Arbeitnehmende nur einen reduzierten Beitrag zum Gesamtsozialversicherungsbeitrag leisten muss. Da Arbeitnehmende mit einem Entgelt in der Gleitzone freiwillig den „normalen“ Sozialversicherungsbeitrag zahlen können, sind nicht automatisch alle Beschäftigungen mit einem entsprechenden Entgelt als Gleitzone gekennzeichnet.

Kategorie	Beschreibung
	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 Ab 01.07.2019: 450,01 bis 1.300,00 Euro Abweichende untere Grenzen existieren für Auszubildende und Beschäftigte in Freiwilligendiensten. Bei mehreren Beschäftigungsverhältnissen ist das insgesamt erzielte Arbeitsentgelt maßgebend. Die entsprechende gesetzliche Regelung besteht seit dem 01.04.2003.

5.5.10 Leiharbeit (leih)

Kategorie	Beschreibung
Variablenlabel	Leiharbeit
Variablenname	leih
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal gibt an, ob es sich bei der Beschäftigung der Person um ein Leiharbeitsverhältnis handelt. 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.
Qualitätshinweis	Im Jahr 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Im Jahr 2012 beträgt die Missing-quote nur noch rund 3% und sinkt bis 2019 auf unter 1%.

5.5.11 Befristung (befrist)

Kategorie	Beschreibung
Variablenlabel	Befristung
Variablenname	befrist
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal gibt an, ob es sich bei der Beschäftigung der Person um ein befristetes oder unbefristetes Arbeitsverhältnis handelt. Das Merkmal entstammt dem Tätigkeitsschlüssel 2010 und ist nur für Meldungen mit einem Enddatum nach dem 30.11.2011 gefüllt.
Qualitätshinweis	Im Jahr 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. Im Jahr 2012 beträgt die Missing-quote nur noch rund 3% und sinkt bis 2019 auf unter 1%.

5.5.12 Abmeldegrund / Abgabegrund / Beendigungsgrund (grund)

Kategorie	Beschreibung
Variablenlabel	Abmeldegrund / Abgabegrund / Beendigungsgrund
Variablenname	grund
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

Kategorie	Beschreibung
Herkunft	BeH, LeH, LHG, ASU, XASU
Datentyp	Numerisch
Detailbeschreibung	**1) BeH** 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 darauf folgenden Jahres-, Unterbrechungs- oder Abmeldung erneut übermittelt werden. 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. **2) LeH** 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. 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. **3) LHG** 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. **4) ASU** 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. 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 **5) XASU**

Kategorie	Beschreibung
	Bei XASU-Sätzen ist der Abmeldegrund enthalten. Im Fall eines Rechtskreiswechsels wird der Satz künstlich gesplittet und als Abmeldegrund "DS durch Splittung erzeugt" eingetragen.
Qualitätshinweis	**1) BeH** 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. **2) LHG, XASU** Der Anteil valider Werte (Füllgrad) des Abgabegrundes ist in der Quelle LHG über alle Jahre sehr niedrig (< 20%).

5.5.13 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 IEB. 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 2005/2006 (siehe Kapitel 4.6) ist ein deutlicher Bruch in den Merkmalsausprägungen zu verzeichnen, da ein einheitliches mapping nicht möglich ist.

5.5.14 Status nach Arbeitsuche (estatnach)

Kategorie	Beschreibung
Variablenlabel	Status nach Arbeitsuche
Variablenname	estatnach
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält für ASU-Sätze den Status der Person nach Abgang aus der Arbeitslosigkeit. Längere Krankheitsepisoden können über dieses Merkmal identifiziert werden. Die Ausprägungen sind folgendermaßen unterteilt:

Kategorie	Beschreibung
	• 1000er: Maßnahme (geförderte Beschäftigung) • 2000er: ungeförderte Beschäftigung • 3000er: in Ausbildung etc. • 4000er: in Selbständigkeit • 5000er: in Sperre • 6000er: in Sonstige
Qualitätshinweis	**1) ASU** Mit dem Umstieg von coArb zu VerBIS 2005/2006 (siehe Kapitel 4.6) ist ein deutlicher Bruch in den Merkmalsausprägungen zu verzeichnen, da ein einheitliches mapping nicht möglich ist. Aber auch sonst zeigen sich je nach Merkmalsausprägung im Zeitverlauf Auffälligkeiten in den relativen Häufigkeiten. Bei Auswertungen ist daher besondere Vorsicht geboten.

5.5.15 Integrationsprognose (ipo)

Kategorie	Beschreibung
Variablenlabel	Integrationsprognose
Variablenname	ipo
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, MTH
Datentyp	numerisch
Detailbeschreibung	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. Die Integrationsprognose "integriert, aber hilfebedürftig" ist nur gültig für Bewerberinnen und Bewerber aus dem Rechtskreis SGB II. Besteht ausschließlich ein Vermittlungswunsch in Ausbildung, ist keine Integrationsprognose festzulegen. 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: 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).
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.

5.5.16 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	Diese Variable ist für Beschäftigungsmeldungen nicht gefüllt! 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.
Qualitätshinweis	Der Anteil der fehlenden Werte ist in dieser Variable besonders hoch und liegt durchschnittlich bei etwa 50%.

5.5.17 Arbeitszeit des Stellengesuchs (arbzeit)

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

5.5.18 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	Je nach Datenquelle hat das Merkmal eine unterschiedliche Bedeutung. **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. Liegt das Enddatum des Leistungsempfangs vor dem 01.01.1998, wird der Restanspruch in Werktagen angegeben. Danach erfolgt die Angabe in Kalendertagen. Die Angabe bezieht sich nicht auf das Enddatum der Episode, sondern auf das Ende des Originalzeitraums.

Kategorie	Beschreibung
	2) MTH Das Merkmal enthält die geplante Dauer der Maßnahme.
Qualitätshinweis	**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. Hierbei handelt es sich vermutlich überwiegend um Korrekturen bei der Bearbeitung von Leistungsfällen. Solche Korrekturen treten zum Beispiel auf, wenn ein Anspruch zu Beginn des Leistungsfalls aufgrund unvollständiger Informationen nicht korrekt festgelegt oder ein noch nicht verfallener Restanspruch nicht berücksichtigt wurde. Diese Korrekturen erfolgen dann aber in der Regel nur in der aktuellen Buchung, eine Bereinigung des kompletten Falls findet nicht statt. Darüber hinaus kann die Angabe zur Anspruchsdauer in Einzelfällen die aus den gesetzlichen Vorgaben abzuleitende individuelle Obergrenze übersteigen. **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.

5.5.19 Trägerart (traeger)

Kategorie	Beschreibung
Variablenlabel	Trägerart
Variablenname	traeger
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LHG, ASU, XASU, MTH, XMTH
Datentyp	Numerisch
Detailbeschreibung	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 Abschnitt 2.2.3. 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.

5.5.20 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

Kategorie	Beschreibung
Detailbeschreibung	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 (begepi) gültig. Folgende Lücken führen nicht zu einer Unterbrechung der Arbeitslosigkeit: • jede Lücke kleiner oder gleich 7 Tage • Krankheitslücken bis zu 42 Tagen

5.5.21 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	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 (begepi) gültig. Folgende Lücken führen nicht zu einer Unterbrechung der Arbeitslosigkeit: • jede Lücke kleiner oder gleich 7 Tage • Krankheitslücken bis zu 42 Tagen 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. Vor 1997 bedeutet der Wert „0“ nicht, dass keine Arbeitslosigkeit vorliegt, da hier die Quellen ASU/XASU nicht gefüllt sind.

5.6 Ortsangaben

5.6.1 Wohnort - Kreis (wo_kreis)

Kategorie	Beschreibung
Variablenlabel	Wohnort - Kreis
Variablenname	wo_kreis
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH, GAV
Datentyp	Numerisch
Hierarchie	Bundesland | Kreis
Detailbeschreibung	**1) BEH, LeH, LHG, XASU, MTH, XMTH** 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 die oder 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

Kategorie	Beschreibung
	unités territoriales statistiques) der Europäischen Union entsprechen die Kreise der Hierarchieebene NUTS 3. 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. Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben für alle Quellen auf den Gebietsstand vom 31.12.2019 umgeschlüsselt, d. h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2019 hatten. Da sich die Grenzen der Kreise im Zeitverlauf geändert haben, würden ohne Gebietsstandsaktualisierung Fälle auftreten, bei denen die Kreiskennziffer des Wohnortes wechselt, ohne dass der oder die betreffende Versicherte umgezogen ist. **2) GAV** Für die Quelle GAV wurde der Wohnortschlüssel aus dem GAV umgeschlüsselt. Dabei wurden aus 301 Kreisen in der DDR im Jahr 1989 durch Gebietsstandreformen 77 Kreise im wiedervereinigten Deutschland im Jahr 2014. Da ein Großteil der Gebietsstandreformen mehrere Kreise von 1989 zu einem Kreis zusammenfügte, ist die Umschlüsselung größtenteils eindeutig. Einige Kreise von 1989 wurden jedoch aufgespalten und mehreren Kreisen im Jahr 2014 zugeordnet. In diesem Fall wurde der Kreis von 1989 dem Kreis von 2014 zugewiesen, der den größten Anteil an Personen aus dem ehemaligen Kreis von 1989 enthält. Somit befindet sich der Gebietsstand für den harmonisierten Wohnort in den GAV-Daten auf dem Stand vom 31.12.2014, da dieser zum Zeitpunkt der Umschlüsselung aktuell war, während die IEB den Gebietsstand vom 31.12.2019 aufweisen. Hierauf ist bei Analysen auf regionaler Ebene zu achten, z. B. indem von Gebietsstandänderungen betroffene Kreise zusammengefasst werden.
Qualitätshinweis	Hinsichtlich der Wohnortangaben in der BeH bestehen für einzelne Beschäftigte Erhebungsungenauigkeiten. Die Meldevorschrift stellt nicht klar, welcher Wohnsitz – Haupt- oder Nebenwohnsitz mit überwiegendem 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. 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 unvergröbert bereitgestellt. Ansonsten wird als Regionalangabe lediglich das Bundesland (wo_bula) bereitgestellt.

5.6.2 Wohnort - Bundesland (wo_bula)

Kategorie	Beschreibung
Variablenlabel	Wohnort - Bundesland
Variablenname	wo_bula
Kategorie	Ortsangaben

Kategorie	Beschreibung
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH, XMTH, GAV
Datentyp	Numerisch
Hierarchie	Bundesland | Kreis
Detailbeschreibung	Dieses Merkmal ist eine Aggregation aus dem Merkmal Wohnort - Kreis (wo_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 in der entsprechenden Variablenbeschreibung.

5.6.3 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 der/des Beschäftigten/Kundin oder 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. Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Angaben für alle Quellen auf den Gebietsstand vom 31.12.2019 umgeschlüsselt, d. h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Agenturbezirk nach den Grenzen, welche die Bezirke zum 31.12.2019 hatten. Da sich die Grenzen im Zeitverlauf geändert haben, würden ohne diese Gebietsstandsaktualisierung Fälle auftreten, bei denen der Bezirk einer Person wechselt, ohne dass diese Person umgezogen ist. Einen Problemfall stellt hinsichtlich der Gebietsstandsaktualisierung allerdings Berlin dar: Die Berliner Arbeitsagenturbezirke wurden im Laufe der Jahre immer wieder neu zugeschnitten, was auch durch die Gebietsstandsumschlüsselung nicht vollständig bereinigt werden konnte. 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. Vor 1999 enthält das Merkmal die Arbeitsagentur, die den Leistungsfall zuletzt bearbeitet hat. Sie wird nicht aus der Wohnortadresse ermittelt und ist nur für die Quelle LeH gefüllt. Außerdem ist eine Gebietsstandsaktualisierung nicht möglich. Dies ist bei Berechnungen im Zeitverlauf (z. B. regionale Arbeitslosenzahlen) zu beachten.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird lediglich das Gebiet der Regionaldirektion, in dem der Wohnort der oder des Versicherten liegt, ausgewiesen.

5.6.4 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.

5.7 Betriebsmerkmale

5.7.1 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)
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WS73 als 3-Steller aus und ist für die Jahre 1975 bis einschließlich 2002 gefüllt. WS73 steht für die "Klassifikation der Wirtschaftszweige 1973“ (Bundesagentur für Arbeit, 2009). Mittels eines dreistelligen Zahlencodes werden 269 Wirtschaftsklassen unterschieden, wobei die erste Stelle dieses Codes die Wirtschaftsabteilung, insgesamt 10, definiert und die beiden ersten Stellen zusammen die jeweilige Gruppe, insgesamt 95, festlegen. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Die Zuordnung zu der entsprechenden Wirtschaftsklasse erfolgt unter Beachtung der institutionellen Ausrichtung des Betriebes.

5.7.2 WZ73 aggregiert (w73_2_agg)

Kategorie	Beschreibung
Variablenlabel	WZ73 aggregiert
Variablenname	w73_2_agg
Kategorie	Betriebsmerkmale
Herkunft	BHP, GAV
Datentyp	Numerisch
Hierarchie	Wirtschaftsabteilung (1-Steller) | Wirtschaftsgruppe (2-Steller)

Kategorie	Beschreibung
Detailbeschreibung	**1) BHP** Dieses Merkmal weist den Wirtschaftszweig gemäß einer aggregierten Klassifikation WS73 als 2-Steller aus und ist für die Quelle BHP für die Jahre 1975 bis einschließlich 2002 gefüllt. Siehe die Beschreibung zum Merkmal WZ73 3-Steller für mehr Informationen zum WS73. **2) GAV** Für die Quelle GAV gibt dieses Merkmal Auskunft über den zuletzt gemeldeten Wirtschaftszweig des Betriebs in der DDR. Diese letzte Meldung erfolgte für mehr als 96% der Betriebe im Jahr 1990, für rund 3% der Betriebe im Jahr 1989 und für weniger als 0,5% der Betriebe im Jahr 1988. Die 722 GAV Wirtschaftszweige werden mithilfe eines Schlüssels von Dietz (1990) auf 269 Wirtschaftszweige des WS73 umgeschlüsselt. Da diese Umschlüsselung nicht immer eindeutig ist, wurden einige Wirtschaftszweige aggregiert, um eine bessere Vergleichbarkeit zu erreichen. In seltenen Fällen ändert sich durch diese Aggregation der 2-Steller des WS73. Um die Robustheit der Ergebnisse zu untersuchen, können Nutzende den Wirtschaftszweig weiter auf die Wirtschaftsabteilung des WS73 aggregieren.

5.7.3 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	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 génerale 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. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w93_3) bereitgestellt.

5.7.4 WZ93 3-Steller (w93_3)

Kategorie	Beschreibung
Variablenlabel	WZ93 3-Steller
Variablenname	w93_3
Kategorie	Betriebsmerkmale

Kategorie	Beschreibung
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller) | Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	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 génerale 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. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.

5.7.5 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	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. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w03_3) bereitgestellt.

5.7.6 WZ03 3-Steller (w03_3)

Kategorie	Beschreibung
Variablenlabel	WZ03 3-Steller
Variablenname	w03_3
Kategorie	Betriebsmerkmale

Kategorie	Beschreibung
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller)| Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	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. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.

5.7.7 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	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. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w08_3) bereitgestellt.

5.7.8 WZ08 3-Steller (w08_3)

Kategorie	Beschreibung
Variablenlabel	WZ08 3-Steller
Variablenname	w08_3
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch

Kategorie	Beschreibung
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller) | Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	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. Jedem Betrieb ist nur eine Kennziffer zugeordnet. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, soll die wirtschaftliche Haupttätigkeit widergespiegelt werden.

5.7.9 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	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. Weitere Informationen zur Klassifikation WS73 finden sich bei der Variablenbeschreibung w73_3.

5.7.10 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	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. Eine ausführliche Beschreibung zur Vorgehensweise findet sich bei Eberle et al., 2011.

5.7.11 w93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen)

Kategorie	Beschreibung
Variablenlabel	w93_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w93_3_gen
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt (1-Steller) | Wirtschaftsabteilung (2-Steller) | Wirtschaftsgruppe (3-Steller) | Wirtschaftsklasse (4-Steller) | Wirtschaftsunterklasse (5-Steller)
Detailbeschreibung	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. Weitere Informationen zur Klassifikation WZ93 finden sich bei der Variablenbeschreibung w93_3.

5.7.12 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	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. Eine ausführliche Beschreibung zur Vorgehensweise findet sich bei Eberle et al., 2011.

5.7.13 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	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

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

5.7.14 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.

5.7.15 Jahr des 1. Auftretens der Betriebsnummer (grd_jahr)

Kategorie	Beschreibung
Variablenlabel	Jahr des 1. Auftretens der Betriebsnummer
Variablenname	grd_jahr
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Bei dieser Variablen handelt es sich um das erste Auftreten der Nummer des Betriebes im Datensatz. Wird eine Betriebsnummer in Westdeutschland erst nach 1975 bzw. in Ostdeutschland nach 1992 erstmalig nachgewiesen, könnte es sich um das Gründungsdatum des betreffenden Betriebes handeln. Allerdings könnte es sich auch um einen bereits länger existierenden Betrieb handeln, dem nach einem 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.

5.7.16 Erstes Auftreten Betriebsnummer (grd_dat)

Kategorie	Beschreibung
Variablenlabel	erstes Auftreten Betriebsnummer
Variablenname	grd_dat
Kategorie	Betriebsmerkmale

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

5.7.17 Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr)

Kategorie	Beschreibung
Variablenlabel	Jahr des letzten Auftretens der Betriebsnummer
Variablenname	lzt_jahr
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	Bei dieser Variablen handelt es sich um das letztmalige Auftreten der Nummer des Betriebes im Datensatz (vgl. Bender et al., 1996). Endet die Existenz einer Betriebsnummer im BHP bereits vor dem Jahr 2008, könnte es sich um eine Betriebsschließung handeln. Allerdings kommen als Ursache auch ein „willkürlicher Wechsel der Betriebsnummer bei Wechsel des Inhabers oder Änderung der Rechtsform“, eine „Ausgliederung von Unternehmensteilen unter einer neuen Nummer“ oder andere verwaltungstechnische Änderungen in Betracht (vgl. Bender et al., 1996 oder Bundesagentur für Arbeit, 2007, S. 9-11). Ein Betrieb muss nicht zwingend in dem Jahr des letzten Auftretens im BHP enthalten sein, da für das BHP nur der Stichtag 30.06. relevant ist. Hat der Betrieb in seinem Schließungsjahr an diesem Stichtag keine Beschäftigte, erscheint er auch nicht im BHP.

5.7.18 Letztes Auftreten Betriebsnummer (lzt_dat)

Kategorie	Beschreibung
Variablenlabel	letztes Auftreten Betriebsnummer
Variablenname	lzt_dat
Kategorie	Betriebsmerkmale
Herkunft	BHP
Datentyp	Datum

Kategorie	Beschreibung
Detailbeschreibung	Bei dieser Variablen handelt es sich um das tagesgenaue letztmalige Auftreten der Nummer des Betriebes im Datensatz (vgl. Bender et al., 1996). Endet die Existenz einer Betriebsnummer im BHP bereits vor dem Jahr 2008, könnte es sich um eine Betriebsschließung handeln. Allerdings kommen als Ursache auch ein „willkürlicher Wechsel der Betriebsnummer bei Wechsel des Inhabers oder Änderung der Rechtsform“, eine „Ausgliederung von Unternehmensteilen unter einer neuen Nummer“ oder andere verwaltungstechnische Änderungen in Betracht (vgl. Bender et al., 1996 oder Bundesagentur für Arbeit, 2007, S. 9-11). Ein Betrieb muss nicht zwingend in dem Jahr des letzten Auftretens im BHP enthalten sein, da für das BHP nur der Stichtag 30.06. relevant ist. Hat der Betrieb in seinem Schließungsjahr an diesem Stichtag keine Beschäftigte, erscheint er auch nicht im BHP.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen tagesgenau bereitgestellt. Ansonsten wird lediglich das Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr) bereitgestellt.

5.7.19 Anzahl Beschäftigte gesamt (az_ges)

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

5.7.20 Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz)

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

5.7.21 Anzahl geringfügig Beschäftigte (az_gf)

Kategorie	Beschreibung
Variablenlabel	Anzahl geringfügig Beschäftigte
Variablenname	az_gf
Kategorie	Generierte Betriebsmerkmale

Kategorie	Beschreibung
Herkunft	BHP
Datentyp	Numerisch
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.

5.7.22 Mittelwert imp. Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)

Kategorie	Beschreibung
Variablenlabel	Mittelwert imp. Bruttotagesentgelt Vollzeitbeschaeftigte
Variablenname	te_imp_mw
Kategorie	Generierte Betriebsmerkmale
Herkunft	BHP
Datentyp	Numerisch
Detailbeschreibung	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. Gemäß den Meldevorschriften zur Sozialversicherung müssen Arbeitgeber angeben, wie hoch das beitragspflichtige Bruttoarbeitsentgelt der 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. 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. 2020). 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.

5.7.23 Arbeitsort - Kreis (ao_kreis)

Kategorie	Beschreibung
Variablenlabel	Arbeitsort Kreis
Variablenname	ao_kreis
Kategorie	Ortsangaben
Herkunft	BHP, GAV
Datentyp	Numerisch
Hierarchie	Bundesland | Kreis
Detailbeschreibung	**1) BHP** Für die Quelle BHP ist jeweils der Kreis (kreisfreie Stadt oder Landkreis) angegeben, in dem der Betrieb der Beschäftigten angesiedelt ist. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des Bundeslandes (NUTS-1), in der ersten bis dritten Stelle den Regierungsbezirk

Kategorie	Beschreibung
	(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. Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben auf den Gebietsstand vom 31.12.2019 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Betriebsortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2019 hatten. Da sich die Grenzen der Kreise im Zeitverlauf geändert haben, würden ohne Gebietsstandsaktualisierung Fälle auftreten, bei denen die Kreiskennziffer des Betriebsortes wechselt, ohne dass dieser Betrieb seinen Standort verlagert hat. **2) GAV** Für die Quelle GAV wurde der Arbeitsortschlüssel aus dem GAV umgeschlüsselt. Dabei wurden aus 301 Kreisen in der DDR im Jahr 1989 durch Gebietsstandsreformen 77 Kreise im wiedervereinigten Deutschland im Jahr 2014. Da ein Großteil der Gebietsstandsreformen mehrere Kreise von 1989 zu einem Kreis zusammenfügte, ist die Umschlüsselung größtenteils eindeutig. Einige Kreise von 1989 wurden jedoch aufgespalten und mehreren Kreisen im Jahr 2014 zugeordnet. In diesem Fall wurde der Kreis von 1989 dem Kreis von 2014 zugewiesen, der den größten Anteil an Personen aus dem ehemaligen Kreis von 1989 enthält. Somit befindet sich der Gebietsstand für den harmonisierten Arbeitsort in den GAV-Daten auf dem Stand vom 31.12.2014, da dieser zum Zeitpunkt der Umschlüsselung aktuell war, während das BHP den Gebietsstand vom 31.12.2019 aufweist. Hierauf ist bei Analysen auf regionaler Ebene zu achten, z. B. indem von Gebietsstandänderungen betroffene Kreise zusammengefasst werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird als Regionalangabe lediglich das Bundesland (ao_bula) bereitgestellt.

5.7.24 Arbeitsort - Bundesland (ao_bula)

Kategorie	Beschreibung
Variablenlabel	Arbeitsort Bundesland
Variablenname	ao_bula
Kategorie	Ortsangaben
Herkunft	BHP, GAV
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 (NUTS-1).

6 Variablenbeschreibung der GAV-spezifischen Merkmale

6.1 Informationen zur Person

6.1.1 Anzahl der Kinder (Kategorien) (GAV_kinder_cat)

Kategorie	**Beschreibung**
Variablenlabel	Anzahl der Kinder (Kategorien)
Variablenname	GAV_kinder_cat
Kategorie	Persönliche Merkmale
Herkunft	GAV
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält die Anzahl der im Haushalt lebenden Kinder bis unter 14 Jahren. 0 Keine Kinder 1 Ein Kind 2 Zwei Kinder 3 Drei Kinder 4 Vier oder mehr Kinder

6.1.2 Schulbildung (GAV_educ)

Kategorie	**Beschreibung**
Variablenlabel	Schulbildung
Variablenname	GAV_educ
Kategorie	Persönliche Merkmale
Herkunft	GAV
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält den Schulabschluss. 1 Ohne Hilfsschulabschluss 2 Hilfsschulabschluss 3 Ohne Klasse 8 4 Klasse 8 5 Klasse 10 6 Abitur mit Berufsausbildung

6.1.3 Universitaetsabschluss (GAV_uniabschl)

Kategorie	**Beschreibung**
Variablenlabel	Universitaetsabschluss
Variablenname	GAV_uniabschl
Kategorie	Persönliche Merkmale
Herkunft	GAV
Datentyp	Numerisch

Kategorie	Beschreibung
Detailbeschreibung	Das Merkmal enthält den Universitätsabschluss. 0 Nein 1 Ja

6.2 Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

6.2.1 Beruf - aktuelle Taetigkeit ist ausbildungsgerecht (GAV_beruf_pass)

Kategorie	Beschreibung
Variablenlabel	Beruf - aktuelle Taetigkeit ist ausbildungsgerecht
Variablenname	GAV_beruf_pass
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	GAV
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal gibt Informationen zu einem ausbildungsgerechten Einsatz in der aktuellen Tätigkeit. Die Merkmale „beruf“ beziehungsweise „beruf_2010_3“ geben die aktuelle Tätigkeit an. 1 Entsprechend dem erworbenen Beruf tätig 2 Dem erworbenen Beruf artverwandt tätig 3 Dem erworbenen Beruf nicht artverwandt tätig 9 Ohne entsprechenden erworbenen Beruf tätig

6.2.2 Beruf - Jahr des Abschlusses, aktueller Taetigkeit (GAV_beruf_jahr)

Kategorie	Beschreibung
Variablenlabel	Beruf - Jahr des Abschlusses, aktueller Taetigkeit
Variablenname	GAV_beruf_jahr
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	GAV
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal gibt Informationen zum Abschlussjahr der aktuellen Tätigkeit. Die Merkmale „beruf“ beziehungsweise „beruf2010_3“ geben die aktuelle Tätigkeit an.
Qualitätshinweis	Einige Werte sind unplausibel hoch oder niedrig. Beispielsweise gibt es Abschlussjahre vor dem Geburtsjahr oder nach 1989.

6.2.3 Beruf - erste erlernte Taetigkeit (KldB 1988), falls mehrere (GAV_beruf_e)

Kategorie	Beschreibung
Variablenlabel	Beruf - erste erlernte Taetigkeit (KldB 1988), falls mehrere
Variablenname	GAV_beruf_e
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	GAV
Datentyp	Numerisch

Kategorie	Beschreibung
Detailbeschreibung	Dieses Merkmal weist den zeitlich ersten Beruf nach KldB 1988 aus, falls eine Person mehr als einen Beruf aufweist. Es basiert auf den Volkswirtschaftlichen Arbeitskräftesystematiken. Zur Vergleichbarkeit mit der IEB Berufsklassifikation wurde für jeden 8-stelligen Berufscode anhand des Vergleiches der Wertelabels ein äquivalenter Berufscode in der KldB 2010 zugewiesen. Dadurch wurden 3.359 GAV Berufscodes 1.268 KldB 2010 Berufscodes zugewiesen. Anschließend wurden diese 1.268 KldB 2010 Berufscodes mit einer Umschlüsselungstabelle der Bundesagentur für Arbeit auf 334 KldB 1988 Berufscodes umgeschlüsselt.
Qualitätshinweis	Für beinahe 15% der Beobachtungen im GAV weist das Merkmal fehlende Werte auf. Fast 75% dieser fehlenden Werte sind bedingt durch fehlende Werte bereits im GAV Berufscode.

6.2.4 Beruf - erste erlernte Taetigkeit (KldB 2010), falls mehrere (GAV_beruf2010_e)

Kategorie	Beschreibung
Variablenlabel	Beruf - erste erlernte Taetigkeit (KldB 2010), falls mehrere
Variablenname	GAV_beruf2010_e
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	GAV
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal weist den zeitlich ersten Beruf nach KldB 2010 aus, falls eine Person mehr als einen Beruf aufweist. Dieses Merkmal basiert auf den Volkswirtschaftlichen Arbeitskräftesystematiken. Zur Vergleichbarkeit mit der IEB Berufsklassifikation wurde für jeden 8-stelligen Berufscode anhand des Vergleiches der Wertelabels ein äquivalenter Berufscode in der KldB 2010 zugewiesen. Dadurch wurden 3.359 GAV Berufscodes 1.268 KldB 2010 Berufscodes zugewiesen. Anschließend wird die 5-stellige KldB 2010 auf den 3-Steller gekürzt.
Qualitätshinweis	Für beinahe 15% der Beobachtungen im GAV weist das Merkmal fehlende Werte auf. Fast 75% dieser fehlenden Werte sind bedingt durch fehlende Werte bereits im GAV Berufscode.

6.2.5 Beruf - Jahr des Abschlusses, erste Taetigkeit, falls mehrere (GAV_beruf_jahr_e)

Kategorie	Beschreibung
Variablenlabel	Beruf - Jahr des Abschlusses, erste Taetigkeit, falls mehrere
Variablenname	GAV_beruf_jahr_e
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	GAV
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal gibt Informationen zum Abschlussjahr der ersten erlernten Tätigkeit. Die Merkmale „GAV_beruf_e“ beziehungsweise „GAV_beruf2010_e“ geben die erste Tätigkeit an.
Qualitätshinweis	Einige Werte sind unplausibel hoch oder niedrig. Beispielsweise gibt es Abschlussjahre vor dem Geburtsjahr oder nach 1989.

6.2.6 Stellung im Beruf (GAV_stib)

Kategorie	Beschreibung
Variablenlabel	Stellung im Beruf

Kategorie	Beschreibung
Variablenname	GAV_stib
Kategorie	Informationen zu Beschäftigung
Herkunft	GAV
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal enthält Informationen zur Stellung im Beruf. 1 Arbeiter und Angestellte 2 Mitglieder von Genossenschaften 4 Lehrlinge

6.2.7 Schichtarbeit in Vollzeit (GAV_schicht_vz)

Kategorie	Beschreibung
Variablenlabel	Schichtarbeit in Vollzeit
Variablenname	GAV_schicht_vz
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	GAV
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal „Schichtarbeit in Vollzeit“ gibt an, ob Beschäftigte in Vollzeit-Schichtarbeit arbeiten.
Qualitätshinweis	Dieses Merkmal ist für Lehrlinge (siehe Variable „GAV_stib“) nicht gefüllt.

6.2.8 Woechentliche Arbeitszeit (GAV_az)

Kategorie	Beschreibung
Variablenlabel	Woechentliche Arbeitszeit
Variablenname	GAV_az
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	GAV
Datentyp	Numerisch
Detailbeschreibung	Das Merkmal enthält Informationen zur wöchentlichen Arbeitszeit.
Qualitätshinweis	In seltenen Fällen haben Teilzeitbeschäftigte eine Arbeitszeit von 40 oder mehr Stunden und Vollzeitbeschäftigte von weniger als 40 Stunden (siehe Variable „teilzeit“). Dieses Merkmal ist für Lehrlinge (siehe Variable „GAV_stib“) nicht gefüllt.

6.2.9 Mehrbetriebsunternehmen (GAV_mehrfbet)

Kategorie	Beschreibung
Variablenlabel	Mehrbetriebsunternehmen
Variablenname	GAV_mehrfbet
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	GAV
Datentyp	Numerisch
Detailbeschreibung	Dieses Merkmal gibt Auskunft darüber, ob die Person in einem Betrieb mit mehreren Teilbetrieben beschäftigt ist.

7 Literatur

Antoni, Manfred (2018): Record linkage of GDR's "Data Fund of Societal Work Power" with administrative labour market biography data of the German Federal Employment Agency. German RLC Working Paper No. wp-grlc-2018-02.

Bender, Stefan/ Hilzendegen, Jürgen/ Rohwer, Götz/ Rudolph, Helmut (1996): Die IAB-Beschäftigtenstichprobe 1975-1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197, Nürnberg.

Berge, Philipp vom/ Frodermann, Corinna/ Schmucker, Alexandra/ Seth, Stefan (2021): Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975 - 2019. FDZ Datenreport, 01/2021 (de), Nürnberg. DOI:10.5164/IAB.FDZD.2101.de.v1

Bertat, Thomas/ Dundler, Agnes/ Grimm, Christopher/ Kiewitt, Jochen/ Schomaker, Christine/ Schridde, Henning/ Zemann, Christian (2013): Neue Erhebungsinhalte 'Arbeitszeit', 'ausgeübte Tätigkeit' sowie 'Schul- und Berufsabschluss' in der Beschäftigungsstatistik. Methodenbericht, Bundesagentur für Arbeit – Statistik, 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: 20.04.2021).

Bundesagentur für Arbeit (Hrsg.) (2005): Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung. Ausgabe Januar 2005, Nürnberg.

Bundesagentur für Arbeit (Hrsg.) (2007): Handbuch für die Betriebsnummernvergabe und –pflege im Rahmen des Meldeverfahren zur Sozialversicherung. Ausgabe Dezember 2007, Nürnberg.

Bundesagentur für Arbeit (2009): Klassifikation der Wirtschaftszweige 1973, Nürnberg. URL: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Wirtschaftszweige/Klassifikation-der-Wirtschaftszweige-1973-2003/Generische-Publikationen/WS-1973.xls?__blob=publicationFile&v=4 (Zugriff: 20.04.2021).

Bundesagentur für Arbeit (2011): Klassifikation der Berufe 2010. Band 1: Systematischer und alphabetischer Teil mit Erläuterungen, Nürnberg. URL: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB-2010/Generische-Publikationen/KldB2010-Printversion-Band1.pdf?__blob=publicationFile&v=7 (Zugriff: 20.04.2021).

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-Die-Klassifikation.xls?__blob=publicationFile&v=8 (Zugriff: 20.04.2021).

Cramer, Ulrich (1985): Probleme der Genauigkeit der Beschäftigtenstatistik. In: Allgemeines Statistisches Archiv 69, S. 56-68.

Dietz, Frido (1990): Umsteigeschlüssel von der Systematik der Volkswirtschaftszweige der vormaligen DDR (Ausgabe 1985) auf das Verzeichnis der Wirtschaftszweige der Bundesanstalt für Arbeit (Ausgabe 1973).

Dietz, Frido/ Helmut Rudolph (1990): Berufstätigenerhebung und der Datenspeicher „Gesellschaftliches Arbeitsvermögen“, Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 4, S. 511-518.

Eberle, Johanna/ Jacobebbinghaus, Peter/ Ludsteck, Johannes/ Witter, Julia (2011): Generation of time-consistent industry codes in the face of classification changes * Simple heuristic based on the Establishment History Panel (BHP). FDZ Methodenreport 05/2011 (en), Nürnberg.

Fitzenberger, Bernd/ Osikominu, Aderonke/ Völter, Robert (2006): Imputation rules to improve the education variable in the IAB employment subsample. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 126, H. 3, S. 405-436.

Fitzenberger, Bernd/ Seidlitz, Arnim (2020): The 2011 break in the part-time indicator and the evolution of wage inequality in Germany. Journal for Labour Market Research 01/2020.

Ganzer, Andreas/ Schmidtlein, Lisa/ Stegmaier, Jens/ Wolter, Stefanie (2021): Betriebs-Historik-Panel 1975-2019. FDZ-Datenreport, 16/2020 v2 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.2016.de.v2.

Gebauer, Ronald/ Dietmar Remy/ Axel Salheiser (2004): Der Datenspeicher „Gesellschaftliches Arbeitsvermögen“: prozessproduzierte Daten als Quelle für die quantitative historische Sozialforschung und eine Soziologie des DDR-Sozialismus. Historical Social Research, 29(4), S. 196-219.

Liepmann, Hannah/ Müller, Dana (2018): A proposed data set for analyzing the labor market trajectories of East Germans around reunification FDZ-Methodenreport 03/2018 (en), Nürnberg.

Ludsteck, Johannes/ Thomsen, Ulrich (2016): Imputation of the Working Time Information for the Employment Register Data. FDZ Methodenreport 01/2016 (en), Nürnberg.

Meinken, Holger/ Koch, Iris (2004): BA-Beschäftigtenpanel 1998-2002. Codebuch, Nürnberg.

Paulus, Wiebke/ Matthes, Britta (2013): Klassifikation der Berufe * Struktur, Codierung und Umsteigeschlüssel. FDZ-Methodenreport 08/2013 (de), Nürnberg.

Rathje, Ulf (1996): Der „Datenspeicher Gesellschaftliches Arbeitsvermögen“ der DDR, Historical Social Research, 21(2), S. 113-117.

Salheiser, Axel (2006): Der Datenspeicher „Gesellschaftliches Arbeitsvermögen“ (DS GAV) des Staatssekretariats für Arbeit und Löhne, in: Best, Heinrich und Dietmar Remy (Eds), Die geplante Gesellschaft, Analysen personenbezogener Datenspeicher der DDR, SFB-580-Mitteilungen, Heft 18, S. 117-128.

Statistisches Bundesamt (1994): Sonderreihe mit Beiträgen für das Gebiet der ehemaligen DDR, Heft 14, Erwerbstätige 1950 bis 1989, Wiesbaden.

Statistisches Bundesamt (2002): Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ 93), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-1993.html (Zugriff: 20.04.2021).

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: 20.04.2021).

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: 02.04.2021).

Statistisches Bundesamt (2021): Staats- und Gebietssystematik, Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluessel_pdf.pdf?__blob=publicationFile&v=5 (Zugriff: 02.04.2021).

Thomsen, Ulrich/ Ludsteck, Johannes/ Schmucker, Alexandra (2018): Skilled or unskilled - Improving the information on qualification for employee data in the IAB Employee Biography. FDZ-Methodenreport 09/2018 (en), Nürnberg.

Wermter, Winfried/ Cramer, Ulrich (1988): Wie hoch war der Beschäftigtenanstieg seit 1983? – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. In: Mitteilungen aus der Arbeitsmarkt – und Berufsforschung 4/88, S. 468-482.

8 Anhang

8.1 Auszählungen

Auszählungen sowie Übersichten über die einzelnen Ausprägungen und Labels der Merkmale finden sich in gesonderten Dateien unter https://fdz.iab.de.

8.2 Abkürzungsverzeichnis

ALG	Arbeitslosengeld
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 –
DIM	Daten- und IT-Management
DÜVO	Zweite VO über die Datenübermittlung auf maschinell verwertbaren Datenträgern im Bereich der Sozialversicherung und der BA – Datenübermittlungs-Verordnung
EDV	Elektronische Datenverarbeitung
FDZ	Forschungsdatenzentrum der Bundesagentur für Arbeit im IAB
FELEG	Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit
GAV	Datenspeicher Gesellschaftliches Arbeitsvermögen
gAw	Träger mit getrennter Aufgabenwahrnehmung
gE	Gemeinsame Einrichtungen
gT	Getrennte Trägerschaft
IAB	Institut für Arbeitsmarkt- und Berufsforschung
IEB	Integrierte Erwerbsbiografien
ISIC	International Standard Industrial Classification of All Economic Activities

LeH	Leistungsempfängerhistorik
LHG	Leistungshistorik Grundsicherung
LVA	Landesversicherungsanstalt
MTH	Maßnahmeteilnahmehistorik
NACE	Nomenclature génerale des activités économiques dans les communautés européennes
NUTS	Nomenclature des unités territoriales statistiques
SAL	Staatssekretariat für Arbeit und Löhne
SGB	Sozialgesetzbuch
SIAB	Stichprobe der Integrierten Arbeitsmarktbiografien
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 13|2022

Veröffentlichungsdatum
12. Dezember 2022

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

Rechte
Nachdruck - auch auszugsweise - nur mit Genehmigung des FDZ gestattet

Bezugsmöglichkeit dieses Dokuments
http://doku.iab.de/fdz/reporte/2022/DR_13-22.pdf

Dokumentation Version
GAV-ADIAB7519_DE_v1_dok1, DOI: 10.5164/IAB.FDZD.2213.de.v1

Datensatz Version
GAV-ADIAB7519_v1, DOI: 10.5164/IAB.GAV-ADIAB7519.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
Matthias Umkehrer
Telefon: 0911 179-6211
E-Mail: Matthias.Umkehrer@iab.de