

Forschungsdatenzentrum

der Bundesagentur für Arbeit
im Institut für Arbeitsmarkt-
und Berufsforschung

FDZ

FDZ-Datenreport

Dokumentation zu Arbeitsmarktdaten

02/2018

DE

NEPS-SC6-Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC6-ADIAB 7515)

Manfred Antoni,
Nadine Bachbauer,
Johanna Eberle,
Basha Vicari



Bundesagentur für Arbeit

NEPS-SC6-Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC6- ADIAB 7515)

Manfred Antoni, Nadine Bachbauer, Johanna Eberle, Basha Vicari
(Institut für Arbeitsmarkt- und Berufsforschung)

Datensatz: NEPS-SC6-ADIAB 7515, Version 1
DOI: 10.5164/IAB.NEPS-SC6-ADIAB7515.de.en.v1

Dokumentation: NEPS-SC6-ADIAB 7515 DE v1 dok1
DOI: 10.5164/IAB.FDZD.1802.de.v1

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	9
1.3	Steckbrief	15
1.4	Merkmalsliste der administrativen Daten	18
1.5	Mengengerüst.....	22
2	DATENQUELLEN	22
2.1	NEPS-SC6-Erhebungsdaten (NEPS-SC6).....	22
2.2	Administrative Daten aus den Integrierten Erwerbsbiografien (IEB)	23
2.3	Beschäftigtenhistorik (BeH)	24
2.4	Leistungsempfängerhistorik (LeH).....	25
2.5	Leistungshistorik Grundsicherung (LHG)	26
2.6	Arbeitsuchendehistorik (ASU und XASU).....	27
2.7	Maßnahmeteilnahmehistoriken (MTH)	27
3	DATENAUFBEREITUNG UND STICHPROBENZIEHUNG	27
3.1	Bereinigungen und Aufbereitungen.....	27
3.2	Episodensplitting	31
3.3	Fehlende Werte	33
4	DATENVERKNÜPFUNG	33
4.1	Informierte Zustimmung der Befragten zur Datenverknüpfung.....	33
4.2	Herkunft der Verknüpfungsmerkmale.....	34
4.3	Bereinigung und Standardisierung personenbezogener Merkmale.....	34
4.4	Exakter Abgleich der Verknüpfungsmerkmale	35
4.5	Fehlertoleranter Abgleich der Verknüpfungsmerkmale	36
4.6	Ergebnis der Datenverknüpfung.....	38
5	DATENQUALITÄT UND PROBLEME	39

5.1	NEPS-SC6-Befragungsdaten (NEPS-SC6).....	39
5.2	Integrierte Erwerbsbiografien (IEB).....	39
5.3	Beschäftigtenhistorik (BeH)	41
5.4	Leistungsempfängerhistorik (LeH).....	43
5.5	Leistungshistorik Grundsicherung (LHG)	43
5.6	Arbeitsuchendenhistorik (ASU/XASU).....	44
5.7	Maßnahmeteilnahmehistoriken (MTH)	45
6	VARIABLENBESCHREIBUNG	46
6.1	Identifikatoren.....	46
6.2	Generierte technische Merkmale	48
6.3	Gültigkeitszeitraum.....	49
6.4	Informationen zur Person	50
6.5	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche.....	55
6.6	Ortsangaben	67
6.7	Betriebsmerkmale	70
6.8	Verknüpfungsmerkmale.....	78
7	LITERATUR	80
8	ANHANG.....	82
8.1	Auszählungen.....	82
9	ABKÜRZUNGSVERZEICHNIS.....	83

Zusammenfassung

Dieser Datenreport beschreibt die verknüpften Erhebungsdaten des Nationalen Bildungspanels der Startkohorte 6 „Bildung im Erwachsenenalter und lebenslanges Lernen“ (NEPS-SC6) mit administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB).

Abstract

This data report describes the linked survey data of Starting Cohort 6 „Adult Education and Lifelong Learning“ of the National Educational Panel Study (NEPS-SC6) with administrative data of the Institute for Employment Research (IAB).

Keywords: German administrative micro data, labor market data, data manual, National Educational Panel Study, survey data

Datenverfügbarkeit

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

1 Einleitung und Kurzbeschreibung

1.1 Einleitung

Dieser Datenreport beschreibt das Datenprodukt „NEPS-SC6-Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC6-ADIAB 7515)“, das gemeinsam vom Leibniz-Institut für Bildungsverläufe e.V. (LifBi) und dem Institut für Arbeitsmarkt- und Berufsforschung (IAB) angeboten wird. Die verknüpften Daten bestehen aus in Interviews erhobenen Informationen von Teilnehmerinnen und Teilnehmern der sogenannten Erwachsenenkohorte (SC6) des Nationalen Bildungspanels (NEPS) und deren Personendaten aus den administrativen Daten des IAB, sofern die Befragten ihre Zustimmung zur Verknüpfung der Daten erteilt haben und in den Daten des IAB identifiziert werden konnten.

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

Inhaltlich zielt die Startkohorte 6 des Nationalen Bildungspanels darauf ab, Bildungs- und Erwerbsverläufe über den Lebensverlauf hinweg nachzuzeichnen sowie Bestimmungsfaktoren formaler, nicht formaler und informeller Bildungsaktivitäten zu ermitteln. Sie ermöglicht die Analyse der Kompetenzausstattung und -entwicklung von Erwachsenen und deren Bedeutung für die Erwerbssituation. Darüber hinaus können mit den Daten die Auswirkungen der Erwerbssituation und der familiären Situation auf Bildungsentscheidungen und die Weiterbildungsbeteiligung untersucht werden, ebenso wie die Erträge formaler Qualifikationen, Kompetenzen und Erwerbserfahrungen. Die Identifikation von Gelegenheiten und Hindernissen für Lernprozesse und Bildung im höheren Erwachsenenalter ist ein weiterer Schwerpunkt der NEPS-SC6-Befragung.

Die Befragung der NEPS-SC6 beinhaltet die Geburtskohorten 1944 bis 1986. Die Stichprobe setzt sich aus drei Teilstichproben zusammen. Aus der IAB-ALWA-Studie², die, ebenso wie

¹ Siehe <https://www.neps-data.de/de-de/datenzentrum/datenunddokumentation/startkohorteerwachsene.aspx>.

² Nähere Informationen zu ALWA finden sich in Kleinert et al. (2011) und zu ALWA-ADIAB in Antoni & Seth (2012). NEPS-SC6-ADIAB ersetzt ALWA-ADIAB, da ein Großteil der ALWA-Studie im NEPS enthalten ist. Ein Vorteil von ALWA-ADIAB ist lediglich, dass in ALWA die vollständige Wohngeschichte seit Geburt erhoben wurde.

die Auffrischungstichprobe, die Geburtsjahrgänge 1956 bis 1986 umfasst, sowie der Aufstockungstichprobe, die die Grundgesamtheit der Studie um die Geburtsjahrgänge 1944 bis 1955 erweitert³. Die Stichprobe basiert auf einer Ziehung aus Einwohnermeldedaten in 240 Gemeinden mit 271 Sample-Points. Für das verknüpfte Datenprodukt steht ausschließlich die Download-Version der NEPS-Daten zur Verfügung. Die Remote- bzw. Onsite-Version, deren Daten weniger stark anonymisiert sind, können für NEPS-SC6-ADIAB nicht bereitgestellt werden.

Die administrativen Daten stammen aus den Integrierten Erwerbsbiografien (IEB) des IAB. Sie enthalten im vorliegenden Fall Daten zu allen Personen, die im Rahmen der NEPS Startkohorte 6 befragt wurden, einer Verknüpfung zugestimmt haben und verknüpft werden konnten, sofern sie im Beobachtungszeitraum mindestens einmal einen der folgenden Zustände aufweisen:

- sozialversicherungspflichtige Beschäftigung in Deutschland (erfasst ab 1975),
- geringfügige Beschäftigung in Deutschland (erfasst ab 1999),
- Bezug von Leistungen nach dem Rechtskreis SGB III (erfasst ab 1975) oder SGB II (erfasst ab 2005),
- bei der Bundesagentur für Arbeit (BA) bzw. den Grundsicherungsträgern als arbeitsuchend gemeldet (erfasst ab 2000),
- Teilnahme an einer arbeitsmarktpolitischen Maßnahme der BA (erfasst ab 2000).

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

Während die Beschäftigtenhistorik (BeH) Ursprung der Informationen zu sozialversicherungspflichtiger und geringfügiger Beschäftigung ist, wird der Bezug von Leistungen nach den Rechtskreisen des SGB III und des SGB II in der Leistungsempfängerhistorik (LeH) bzw. in der Leistungshistorik Grundsicherung (LHG) erfasst. Die Arbeitsuchendenhistoriken (ASU und XASU) stellen die Datenquelle für die durch die BA bzw. die durch die kommunalen Grundsicherungsträger erfassten Zeiten der Arbeitsuche dar, während Teilnahmen an arbeitsmarktpolitischen Maßnahmen in der Maßnahmeteilnahmehistorik (MTH) erfasst werden.

³ Die erste Welle der Startkohorte 6 des NEPS setzt sich aus Informationen aus der ALWA-Studie zusammen, während die zweite Welle aus Informationen der ersten Haupterhebung im Rahmen der NEPS-Studie besteht. Im Folgenden wird die Wellen-Logik verwendet, die auch in den Daten des NEPS Anwendung findet. In Abgrenzung dazu findet sich in manchen Dokumentationsmaterialien des NEPS die Kennzeichnung von NEPS-Haupterhebungen (ohne Berücksichtigung der ALWA-Studie).

Neben diesen personenbezogenen Daten können auch Angaben zu den beschäftigenden Betrieben aus dem Betriebs-Historik-Panel (BHP) des IAB (Schmucker et al., 2016) zugespielt werden.

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

1.2 Datennutzung

1.2.1 Datenzugang und Datenhaltung

Die NEPS-SC6-ADIAB-Daten sind schwach anonymisiert und können ausschließlich im Rahmen von Gastaufenthalten an einem der Standorte des Forschungsdatenzentrums der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung und anschließender Datenfernverarbeitung ausgewertet werden.

Für die Datennutzung ist in einem ersten Schritt ein Antrag beim IAB-FDZ⁴ einzureichen, über die Genehmigung des Forschungsprojekts entscheidet das Bundesministerium für Arbeit und Soziales (BMAS). Nach erfolgter Genehmigung muss ein Datennutzungsvertrag für die NEPS-Daten mit dem Forschungsdatenzentrum des LfBi geschlossen werden⁵. Abschließend wird, sobald der Datennutzungsvertrag für die NEPS-Daten dem IAB vorliegt, ein Nutzungsvertrag zwischen der Institution der Forscherin bzw. des Forschers und dem IAB geschlossen. Details zur Beantragung des Datensatzes und den Datenverarbeitungsmöglichkeiten können der Website des IAB-FDZ entnommen werden. Der Datenzugriff beginnt somit erst, wenn die antragstellende Person nachweislich Verträge mit beiden Forschungsdatenzentren abgeschlossen hat.

Zur Beantwortung von Fragen zu diesem Datenprodukt können sich Nutzerinnen und Nutzer an die Mitarbeiterinnen und Mitarbeiter der FDZs am IAB und am LfBi wenden. Die Arbeitsteilung der beiden Einrichtungen sieht vor, dass sich die Nutzerinnen und Nutzer mit Fragen zu den NEPS-Daten ausschließlich an das LfBi-FDZ wenden, mit Fragen zu den administrativen Daten, zu technischen Aspekten der Datenverknüpfung oder des Zugriffs auf die Daten

⁴ Um die beiden beteiligten Forschungsdatenzentren leicht unterscheiden zu können, wird das FDZ der BA im IAB im Folgenden als IAB-FDZ bezeichnet, das Forschungsdatenzentrum am LfBi dagegen als LfBi-FDZ.

⁵ <https://www.neps-data.de>

ausschließlich an das IAB-FDZ. Bei allgemeinen Fragen zum NEPS kann neben dem IAB-FDZ auch das NEPS-Forum⁶ konsultiert werden.

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

Ein weiteres Element von NEPS-SC6-ADIAB ist die Verknüpfungsdatei, welche die Personennummer sowie technische Merkmale mit Informationen darüber beinhaltet, ob bzw. wie die Verknüpfung mit den administrativen Daten gelungen ist. Die Datei umfasst auch die Befragten der NEPS Startkohorte 6, zu denen keine administrativen Daten gefunden werden konnten. Somit können Ausfallanalysen auf Basis aller Teilnehmerinnen und Teilnehmer der NEPS-SC6 durchgeführt werden.

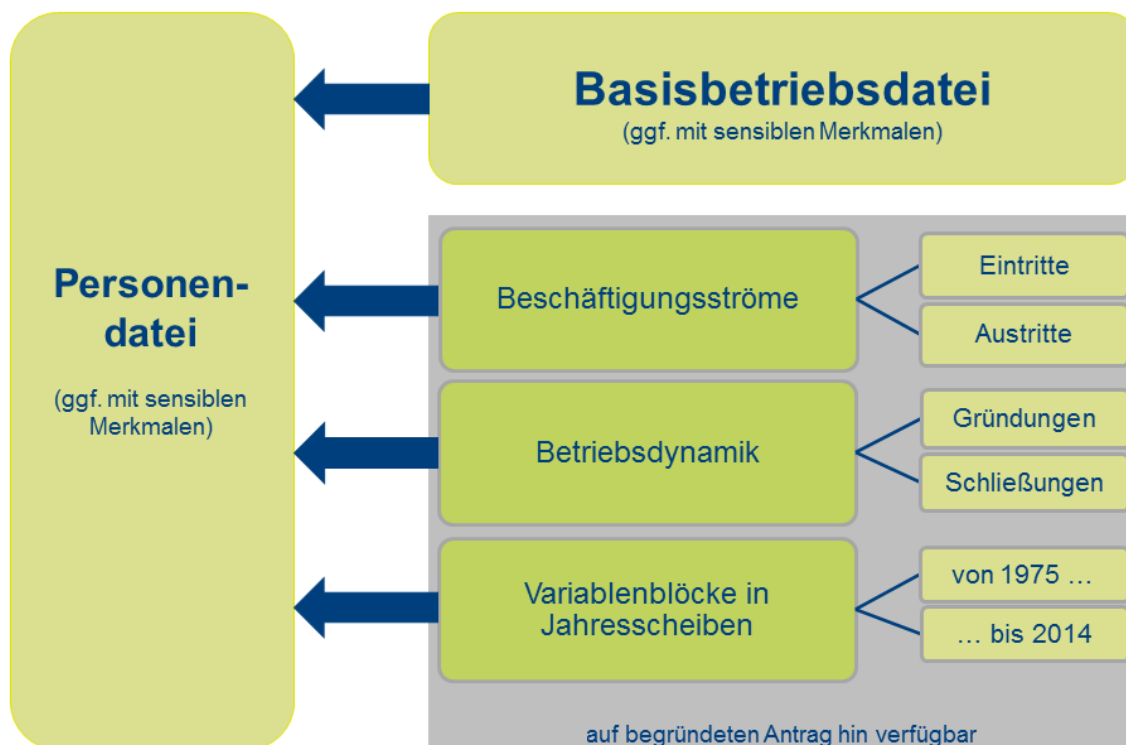
Die administrativen Daten, die analog zu den NEPS-Daten, sowohl deutsche als auch englische Labels enthalten,⁷ sind modular gegliedert und in mehreren Dateien abgelegt. Ein Modul, das im weiteren Verlauf „Personendatei“ genannt wird, enthält Identifikatoren (systemfreie Personen- und Betriebsnummer), persönliche Merkmale, Informationen zu Beschäftigung, Leistungsbezug, Arbeitsuche und Teilnahme an arbeitsmarktpolitischen Maßnahmen, Merkmale zum Wohnort sowie technische Merkmale. Diese Personendaten umfassen tagesgenaue Angaben für den Zeitraum 1975 bis 2014.

Ein zweites Modul, die Basisbetriebsdatei, enthält die Betriebsnummer, das Jahr, Angaben zu Arbeitsort und Wirtschaftszweigen sowie weitere Betriebsinformationen zum Stichtag 30. Juni der Jahre 1975 bis 2014. Für Ostdeutschland liegen die Betriebsdaten erst ab 1992 vor. Somit sind administrative Betriebs- und Personenmerkmale in getrennten Dateien abgelegt (vgl. Abbildung 1). Zusätzlich sind in der Basisbetriebsdatei generierte Variablen für über die Zeit konsistente Wirtschaftszweigklassifikationen enthalten (s. Eberle et al., 2011). Zudem können weitere Betriebsvariablen und Erweiterungsmodule des BHP genutzt werden, die von Schmucker et al. (2016) beschrieben werden. Die Verknüpfung der administrativen Personen- und Betriebsdateien erfolgt über die Betriebsnummer und das Jahr des Datensatzes. Die in der Merkmalsliste (vgl. S. 18ff.) mit einem „*“ gekennzeichneten Variablen sind in der Basisbetriebsdatei enthalten.

⁶ <https://forum.neps-data.de/>

⁷ Mit den Stata-Befehlen *label language en* bzw. *label language de* kann auf die englisch- bzw. deutschsprachigen Labels umgestellt werden.

Abbildung 1 Datenhaltung der administrativen Daten



1.2.2 Sensible Merkmale und Zusatzvariablen

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

Administrative Personendatei:

- Staatsangehörigkeit (nation)
- Wohnort Arbeitsagentur (wo_aa)
- Wohnort Kreis (wo_kreis)
- Berufsuntergruppe (beruf2010_4)

Basisbetriebsdatei:

- Arbeitsort Kreis (ao_kreis)
- Wirtschaftszweig - Wirtschaftsunterklasse (w93_5)
- Wirtschaftszweig - Wirtschaftsunterklasse (w03_5)
- Wirtschaftszweig - Wirtschaftsunterklasse (w08_5)
- Erstes Auftreten Betriebsnummer (grd_dat)
- Letztes Auftreten Betriebsnummer (lzt_dat)

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

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

Auch die NEPS-Daten liegen in unterschiedlichen Anonymisierungsgraden vor. Bei den für NEPS-SC6-ADIAB verfügbaren Daten handelt es sich um die Download-Version, welche den höchsten Anonymisierungsgrad aufweist. Diese Version enthält einige Variablen, jeweils im Namen mit den Suffixen „_R“ bzw. „_O“ gekennzeichnet, deren Ausprägungen durch Aggregation oder andere Maßnahmen anonymisiert wurden. Die vollständigen Informationen dieser Variablen sind nur in der Remote- bzw. Onsite-Version verfügbar. Aus Datenschutzgründen können weder die Remote-Version noch die Onsite-Version für NEPS-SC6-ADIAB bereitgestellt werden, auch nicht durch einen gesonderten Antrag. Es besteht somit keine Möglichkeit, Variablen aus der NEPS-Erhebung zu nutzen, die nicht in der Download-Version enthalten sind. Dies betrifft vornehmlich Regionaldaten wie Regierungsbezirke und Kreise sowie Länderkennzeichnungen wie Muttersprache oder Herkunftsland.

1.2.3 Dateinamen der administrativen Daten und der Verknüpfungsdatei

Personendatei

NEPS-SC6-ADIAB_7515_v1.dta

Basisbetriebsdatei

NEPS-SC6-ADIAB_7515_v1_bhp_basis_v1.dta

Verknüpfungsdatei

NEPS-SC6-ADIAB_7515_v1_schluessel.dta

Zusatzmodule BHP

Bestandsdaten mit Variablenblöcken

NEPS-SC6-ADIAB_7515_v1_bhp_v1_####.dta, #### = 1975 - 2014

Beschäftigtenströme

NEPS-SC6-ADIAB_7515_v1_bhp_inflow_v1.dta

NEPS-SC6-ADIAB_7515_v1_bhp_outflow_v1.dta

Betriebsdynamik

NEPS-SC6-ADIAB_7515_v1_bhp_entry_v1.dta

NEPS-SC6-ADIAB_7515_v1_bhp_exit_v1.dta

1.2.4 Dateinamen der NEPS-Startkohorte 6

Tabelle 1 Datenstruktur NEPS-SC6

Datensatz	Dateiname	Inhalt und Besonderheiten
Basisdaten	Basics.dta	Informationen zu aktuellen Basismerkmalen der Befragten (z. B. Soziodemographie, aktuelle Erwerbstätigkeits- und Haushaltscharakteristika); Querschnittsdaten im Wide-Format
Profildaten	CohortProfile.dta	Informationen zur Teilnahme, zum Interviewdatum, zur Verfügbarkeit von Befragungs- und Kompetenzdaten; Ausgangsdatei für das Zusammenspielen von Informationen (Identifikatoren); Längsschnittdaten im Long-Format
Methodendaten zu Befragungen	Methods.dta	Informationen zum Feldverlauf und zur Datenerhebung (Rücklauf, Kontaktierung, Dauer, Interviewer, Incentives); Längsschnittdaten im Long-Format
Gewichtungsdaten	Weights.dta	Informationen zu Stichprobencharakteristika und Stratifizierung inklusive Gewichtungsvariablen; Längsschnittdaten im Long-Format
Befragungsdaten	pTarget.dta	Informationen aus den Befragungen der Zielperson; Längsschnittdaten im Long-Format
Kompetenzdaten	xTarget Competencies.dta	Informationen zu den in Welle 3, 5 und 7 durchgeführten Kompetenztestungen; Längsschnittdaten im Wide-Format
Methodendaten zu Kompetenztestungen	Methods Competencies.dta	Informationen zur methodischen Durchführung der Kompetenztestungen; Längsschnittdaten im Long-Format
Lebensverlaufs-Spell-daten	Biography.dta	Integrierte und geglättete Lebensverlaufsdaten aus verschiedenen Lebensbereichen; Zusammenfassung aller vollständigen, harmonisierten und rechtszentrierten Episoden-Spells
Familienstand-Spell-daten	MaritalStates.dta	Informationen zu Ereignissen bzgl. Familienstand (verheiratet, geschieden und verwitwet); aus spPartner generierter Datensatz
Kinderdaten	Children.dta	Basisinformationen zu Kindern, mit denen die Zielperson je zusammengelebt hat; aus spChild generierter Datensatz
Schul- und Ausbildungsdaten	Education.dta	Informationen zu den Übergängen in der (Aus-)Bildungskarriere der Zielperson (z. B. CASMIN, ISCED-97); aus den Spell-Datensätzen zur Schul- und Ausbildung generierter Datensatz
Aus- und Weiterbildungsdaten	FurtherEducation.dta	Integrierte Informationen zu allen Kursdaten, auch retrospektiv; aus den Spell-Datensätzen zu Aus- und Weiterbildungen generierter Datensatz
Schulbildungs-Spell-daten	spSchool.dta	Informationen zur Allgemeinbildungshistorie, von der Einschulung bis zum (antizipierten) Zeitpunkt des Abschlusses; Regionaldaten (Land, Bundesland)
Berufsvorbereitung-Spell-daten	spVocPrep.dta	Informationen zu Episoden der Berufsvorbereitung nach der allgemeinen Schulbildung
Ausbildung-Spell-daten	spVocTrain.dta	Informationen zu allen weiterführenden Ausbildungen, die die Zielperson jemals absolviert hat (berufliche

Datensatz	Dateiname	Inhalt und Besonderheiten
		Ausbildung, Studium); Regionaldaten (Land, Bundesland)
Erwerbstätigkeit-Spell- daten	spEmp.dta	Informationen zu Episoden regulärer Erwerbstätigkeiten inklusive Praktika; Beginn einer neuen Episode bei Arbeitgeber-, Tätigkeitswechsel oder Unterbrechung der Erwerbstätigkeit; Regionaldaten (Land/Bundesland)
Erwerbslosigkeit-Spell- daten	spUnemp.dta	Informationen zu Episoden von Erwerbslosigkeit unabhängig von der Registrierung als arbeitslos
Wehr- und Zivildienst- Spelldaten	spMilitary.dta	Informationen zu Episoden von Wehr- und Zivildienst als auch Zeiträume, die für Freiwilligenarbeit im sozialen Bereich oder dem Umweltsektor genutzt wurden
Partner-Spelldaten	spPartner.dta	Informationen zur Historik von Partnerschaften der Zielperson (Zusammenleben, Heirat, Trennung, Scheidung); Merkmale des Partners (z. B. Geburtsdatum, Ausbildung, Beruf)
Kinder-Spelldaten	spChild.dta	Informationen zu biologischen, Pflege- und Adoptivkindern der Zielperson sowie zu allen Kindern, die mit der Zielperson zusammenleben oder gelebt haben
Elternzeiten-Spelldaten	spParLeave.dta	Information zu jedem Kind aus spChild (ausgenommen verstorbene Kinder) und Information, ob Zielperson Elternzeit genommen hat
Zusammenleben mit Kindern-Spelldaten	spChildCohab.dta	Informationen zur Historik des Zusammenlebens der Zielperson mit Kindern
Lücken-Spelldaten	spGap.dta	Informationen zu Lücken in der Erfassung des individuellen Lebensverlaufs; identifiziert durch Prüfmodul
(Berufliche) Kurse- Spelldaten	spCourses.dta	Informationen zu den innerhalb der letzten 12 Monate besuchten Kursen und Weiterbildungen während einer Erwerbstätigkeits-, Erwerbslosigkeits-, Elternzeit-, Wehr-/Zivildienst- oder Lückenepisode
Weitere Kurse-Spell- daten	spFurtherEdu1.dta	Informationen zu Kursen, die neben den in spCourses und spVocTrain berichteten Kursen absolviert wurden; sowohl berufliche Weiterbildungen als auch Kurse mit privaten Zielen (z. B. Yoga, Kochkurse)
Zusätzliche Kursinfo- Spelldaten	spFurtherEdu2.dta	Zusätzlich erhobene Informationen zu zwei zufällig ausgewählten Kursen aus spVocTrain, spCourses und spFurtherEdu1 (z. B. Kosten, Motivation, Zertifikate)
Deutschkurse-Spell- daten	spFurtherEdu3.dta	Informationen zu Kursen mit Deutsch als Fremdsprache; nur für Migrantinnen und Migranten erhoben
Schulabschlüsse im Ausland-Spelldaten	spSchoolExt Exam.dta	Informationen zu im Ausland erworbenen Schulabschlusszertifikaten
Ausbildungsabschlüsse im Ausland-Spelldaten	spVocExtExam.dta	Informationen zu im Ausland erworbenen Ausbildungsabschlusszertifikaten
Ehrenamt-Spelldaten	spVolunteer- Work.dta	Informationen zu ehrenamtlichen Tätigkeiten (z. B. Dauer, Anforderungen, Inhalt der Tätigkeit)
Wohnort-Spelldaten	spResidence.dta	Informationen zur Wohnbiographie (nur für ursprüngliche ALWA-Befragte); Regionaldaten (Land, Bundesland)

Welche Variablen in den jeweiligen Datensätzen als eindeutige Identifikatoren zum Zusammenspielen von Informationen fungieren, kann der Merging-Matrix auf der NEPS-Website entnommen werden. Weitere Informationen zu den einzelnen Datensätzen finden sich im Data Manual der SC6, welches ebenfalls von der Webseite bezogen werden kann.

1.3 Steckbrief

Tabelle 2 Steckbrief NEPS-SC6-ADIAB

Themen/ Merkmalsgruppen	<p><u>Erhebungsdaten</u> Längsschnittdaten zu Soziodemographie; Bildungs- und Erwerbsbiographie; Kompetenzentwicklungen, Bildungsprozesse, Bildungsentscheidungen und Bildungsrenditen in formalen, nicht formalen und informellen Kontexten</p> <p><u>Administrative Personendaten</u></p> <p>Beschäftigtenhistorik (BeH): Jahres- und Abmeldungen zur Sozialversicherung für sozialversicherungspflichtig und geringfügig Beschäftigte</p> <p>Leistungsempfängerhistorik (LeH): Informationen zu Leistungsbezug nach SGB III für Bezieher von Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld</p> <p>Leistungshistorik Grundsicherung aus A2LL (LHG): Daten zu Personen, die Leistungen zur Sicherung des Lebensunterhalts nach SGB II beziehen (Trägerarten: ARGE/gemeinsame Einrichtungen, getrennte Trägerschaft/getrennte Aufgabenwahrnehmung, zugelassene kommunale Träger)</p> <p>Arbeitsuchendehistorik (ASU): Informationen zur Arbeitsuche</p> <p>Arbeitsuchendehistorik aus XSozial-BA-SGB II (XASU): Informationen zur Arbeitsuche, die über den Übermittlungsstandard XSozial-BA-SGB II von zugelassenen kommunalen Trägern an die BA gemeldet werden.</p> <p>Maßnahmeteilnahmehistoriken (MTH): Informationen zu (geplante)Teilnahmen an Maßnahme der aktiven Arbeitsmarktförderung (keine Maßnahmen der zugelassenen kommunalen Träger im SGB II)</p> <p><u>Administrative Betriebsdaten</u></p> <p>Kerndatensatz: Betriebsmerkmale (z.B. systemfreie Betriebsnummer, Wirtschaftszweig, Bundesland), Erwerbsstruktur der Beschäftigten (z.B. Anzahl der Beschäftigten nach Geschlecht, Art der Beschäftigung), Altersstruktur der Beschäftigten, Struktur der Beschäftigten nach Schul- und Berufsausbildung, Struktur der Beschäftigten nach Berufsklassifikation (Blossfeld Klassifikation), Durchschnittslöhne der Vollzeitbeschäftigten</p> <p>Erweiterungsmodul- Beschäftigtenströme: Anzahl Zugänge gesamt, Struktur der Zugänge nach Geschlecht, Beschäftigungsart, Berufsklassifikation (Blossfeld Klassifikation), Wiedereinstellung, Alter, Anzahl Abgänge gesamt, Struktur der Abgänge nach Geschlecht, Beschäftigungsart, Berufsklassifikation (Blossfeld Klassifikation), Betriebszugehörigkeit, Alter</p> <p>Erweiterungsmodul- Betriebsdynamik: Gründungsjahr, Art der Gründung, Hilfsvariablen zur Klassifizierung der Gründung, Schließungsjahr, Art der Schließung, Hilfsvariablen zur Klassifizierung der Schließung</p>
-------------------------	---

Untersuchungseinheit	<u>Erhebungsdaten:</u> Personen der Wohnbevölkerung Deutschlands (zum Zeitpunkt der Stichprobenziehung) im Erwerbsalter unabhängig von Erwerbsstatus, Nationalität und Deutschkenntnissen <u>Administrative Personendaten:</u> Sozialversicherungspflichtig Beschäftigte (ab 1999 auch geringfügig Beschäftigte), Leistungsempfänger, Arbeitssuchende, Maßnahmeteilnehmerinnen und -teilnehmer
Fallzahlen	<u>Erhebungsdaten:</u> Welle 1 (ALWA): 6.778 Teilnehmerinnen und Teilnehmer (davon verknüpft: 5.354 Personen) Welle 2: 11.649 Teilnehmerinnen und Teilnehmer (davon zusätzlich verknüpft: 3.416) Welle 3: 9.336 Teilnehmerinnen und Teilnehmer Welle 4: 14.112 Teilnehmerinnen und Teilnehmer (davon zusätzlich verknüpft: 3.888) Welle 5: 11.696 Teilnehmerinnen und Teilnehmer Welle 6: 10.639 Teilnehmerinnen und Teilnehmer Welle 7: 9.770 Teilnehmerinnen und Teilnehmer Welle 8: 9.236 Teilnehmerinnen und Teilnehmer <u>Administrative Daten:</u> 12.660 Personen 454.974 Originalsätze 599.833 überschneidungsfreie Sätze (nach Episodensplitting)
Zeitraum	<u>Erhebungsdaten:</u> Welle 1 (ALWA): 2007/2008 Welle 2: 2009/2010 Welle 3: 2010/2011 Welle 4: 2011/2012 Welle 5: 2012/2013 Welle 6: 2013 Welle 7: 2014/2015 Welle 8: 2015/2016 <u>Administrative Daten:</u> Der Zeitraum hängt von der Datenquelle ab. BeH: 01.01.1975 - 31.12.2014 (2012: 30-Monats-Datei, 2013: 18-Monats-Datei; 2014: 6-Monats-Datei) LeH: 29.11.1976 - 31.12.2014 ASU: 07.11.1994 - 31.12.2014 LHG: 01.01.2005 - 31.12.2014 XASU: 01.01.2005 - 31.12.2014 MTH: 02.11.1993 - 31.12.2014
Zeitbezug	<u>Erhebungsdaten:</u> Befragungszeitpunkt; Spell- bzw. Ereignisdaten retrospektiv bei Erstbefragung, anschließend jeweils seit letzter Befragung <u>Administrative Personendaten:</u> Tagesgenaue Erwerbsbiografien <u>Administrative Betriebsdaten:</u> Stichtag 30. Juni
Regionale Gliederung	Bundesländer, Kreise

Gebietsstand	Gebietsstand bereinigt zum 31.12.2014
Erhebungsdesign	<p><u>Erhebungsdaten:</u></p> <p>Welle 1: ALWA (Arbeiten und Lernen im Wandel)</p> <p>Welle 2: Wiederholungsbefragung ALWA, Auffrischung sowie Aufstockung des Panels, CATI-CAPI biografisches Interview</p> <p>Welle 3: Wiederholungsbefragung Vorwelle, CATI-CAPI biografisches Interview und PAPI Kompetenztestung (Lesen/Mathematik/Metakognition)</p> <p>Welle 4: Wiederholungsbefragung Vorwelle und Auffrischung, CATI-CAPI biografisches Interview</p> <p>Welle 5: Wiederholungsbefragung Vorwelle, CATI-CAPI biografisches Interview und PAPI Kompetenztestung (Lesen/Naturwissenschaften/ICT/Metakognition)</p> <p>Welle 6: Wiederholungsbefragung Vorwelle, CATI-CAPI biografisches Interview</p> <p>Welle 7: Wiederholungsbefragung Vorwelle, CATI-CAPI biografisches Interview und PAPI Kompetenztestung (Kognitive Grundfähigkeiten/Hörverstehen/Metakognition)</p> <p>Welle 8: Wiederholungsbefragung Vorwelle, CATI-CAPI biografisches Interview</p> <p><u>Administrative Daten:</u></p> <p>Identifikation der Befragten in den Integrierten Erwerbsbiografien des IAB</p>
Beteiligte Institutionen	<p><u>Erhebungsdaten:</u></p> <p>Auftraggeber: Leibniz-Institut für Bildungsverläufe e.V. (IIfBi) an der Otto-Friedrich-Universität Bamberg</p> <p>Durchführung der Befragung: infas Institut für angewandte Sozialwissenschaft GmbH</p> <p><u>Administrative Daten:</u></p> <p>Sozialversicherungsträger, Bundesagentur für Arbeit</p>
Frequenz der Datensammlung	<p><u>Erhebungsdaten</u></p> <p>ca. jährlich</p> <p><u>Administrative Daten</u></p> <p>Laufend</p>
Dateiformat/ -größe	<p><u>Erhebungsdaten</u></p> <p>Stata; 31 Einzeldatensätze zw. 66 KB und 143,1 MB</p> <p><u>Administrative Daten</u></p> <p>Stata; Personendatensatz: 42,4 MB; Betriebsdatensatz: 1,65 MB</p>
Dateiorganisation	<p><u>Erhebungsdaten</u></p> <p>Die Daten sind in mehreren Dateien abgelegt, die im Quer- oder Längsschnitt (wellen- bzw. episodenspezifisch) angelegte Personendaten enthalten.</p> <p><u>Administrative Daten</u></p> <p>Die Daten werden in zwei Dateien abgelegt. Die eine enthält personenbezogene und die andere betriebsbezogene Informationen. Auf begründeten Antrag werden weitere Dateien mit zusätzlichen Betriebsinformationen bereitgestellt.</p> <p><u>Verknüpfungsdaten</u></p> <p>Daneben gibt es eine Datei, die die Identifikationsvariable sowie Informationen zum Matchingerfolg beinhaltet.</p>
Datenzugang	Gastaufenthalt am IAB-FDZ, Datenfernverarbeitung beim IAB-FDZ nach vorangegangenem Gastaufenthalt
Anonymisierungsgrad	Schwach anonymisiert

Sensible Merkmale	Staatsangehörigkeit (nation), Berufsuntergruppe (beruf2010_4), Wohnort – Kreis (wo_kreis), Wohnort – Arbeitsagentur (wo_aa), Arbeitsort – Kreis (ao_kreis), Wirtschaftsunterklasse (w93_5), Wirtschaftsunterklasse (w03_5), Wirtschaftsunterklasse (w08_5), Erstes Auftreten der Betriebsnummer (grd_dat), Letztes Auftreten der Betriebsnummer (lzt_dat)
Zitierung der Daten und Datendokumentation	<p><u>Daten:</u> „Die Datengrundlage dieses Beitrags bilden die NEPS-SC6-Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC6-ADIAB). 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.NEPS-SC6-ADIAB7515.de.en.v1</p> <p><u>Datendokumentation:</u> Antoni, Manfred; Bachbauer, Nadine; Eberle Johanna, Vicari, Basha (2018): NEPS-SC6-Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC6-ADIAB 7515). FDZ-Datenreport, 02/2018 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.1802.en.v1</p>
Datensatzversion	NEPS-SC6 Erhebungsdaten verknüpft mit administrativen Daten des IAB (NEPS-SC6-ADIAB 7515); DOI: 10.5164/IAB.NEPS-SC6-ADIAB7515.de.en.v1

1.4 Merkmalsliste der administrativen Daten

Die Merkmalsübersicht in Tabelle 4 listet die Variablennamen und die längeren Merkmalsbezeichnungen der administrativen Daten auf. Seitens der Betriebsdaten sind hier wie in Kapitel 6 nur die Merkmale der Basisbetriebsdatei aufgeführt. Die weiteren Betriebsmerkmale sowie die Erweiterungsmodule werden ausführlich in Schmucker et al. (2016) beschrieben.

Weiterhin bietet Tabelle 4 eine Übersicht darüber, welche Merkmale für welche Datenquellen gefüllt sind, wobei Tabelle 3 aufzeigt, wie die Füllgrade zu interpretieren sind.

Tabelle 3 Füllgrad der Merkmale

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

Beispiel: Das Merkmal Tagesentgelt bzw. täglicher Leistungssatz ist nur für BeH- und LeH-Sätze gefüllt. Die Sätze der anderen Datenquellen enthalten in dieser Variable den Missing-Wert „.n“. Eine weitere Besonderheit liegt darin, dass einige Merkmale je nach Datenquelle unterschiedliche Inhalte haben. So enthält der Erwerbsstatus für BeH-Sätze die Personengruppe des Meldeverfahrens, für LeH-Sätze die Leistungsart, für LHG-Sätze den Status

SGB II, für ASU- bzw. XASU-Sätze den Arbeitsuchestatus und für MTH die Maßnahmenkategorie. Nicht bei jedem Merkmal sind diese Unterschiede direkt aus dem Variablennamen ersichtlich.

Tabelle 4 Merkmalsliste der administrativen Daten mit Füllgrad

Merkmalsliste	BHP Basis	Seite	Beh	LeH	LHG	ASU	XASU	MTH
Identifikatoren		46						
Personen-ID für NEPS-SC6-ADIAB (ID_t)		46						
Betriebs-ID für NEPS-SC6-ADIAB (betnr)	*	46						
Generierte technische Merkmale		48						
Satzzähler pro Konto (spell)		48						
Quelle des Satzes (quelle)		48						
Jahr (jahr)	*	48						
Gültigkeitszeitraum		49						
Beginndatum Originalbeobachtung (begorig)		49						
Enddatum Originalbeobachtung (endorig)		49						
Beginndatum der Episode (begepi)		50						
Enddatum der Episode (endepi)		50						
Informationen zur Person		50						
Geschlecht (frau)		50						
Geburtsjahr (gebjahr)		50						
Staatsangehörigkeit (nation)		51						
Staatsangehörigkeit vergrößert (nation_gr)		51						
Familienstand (famst)		51						
Kinderzahl (kind)		52						
Ausbildung (ausbildung)		52						
Schulbildung (schule)		54						
Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche		55						

Merkmalsliste	BHP Basis	Seite	BeH	LeH	LHG	ASU	XASU	MTH
Tagesentgelt/ täglicher Leistungssatz (tentgelt)		55						
Beruf – ausgeübte/letzte Tätigkeit (KldB 1988) (beruf)		56						
Berufsgruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_3)		57						
Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KldB 2010) (beruf2010_4)		57						
Anforderungsniveau – ausgeübte/letzte Tätigkeit (KldB 2010) (niveau)		58						
Teilzeit (teilzeit)		59						
Erwerbsstatus (erwstat)		59						
Gleitzone (gleitz)		61						
Leiharbeit (leih)		62						
Befristung (befrist)		62						
1.1.1 Grund Abgabe Beschäftigungsmeldung/Grund Ende Leistungsbezug (grund)		62						
Erwerbsstatus vor Arbeitsuche (estatvor)		63						
Status nach Arbeitsuche (estatnach)		64						
Profillage (profil)		64						
Art der Kündigung der letzten Tätigkeit (art_kuend)		65						
Arbeitszeit des Stellengesuchs (arbeitszeit)		65						
Restanspruch/geplante Dauer (restanspruch)		66						
Trägerart (traeger)		66						
Beginndatum der Arbeitslosigkeit (alo_beg)		67						
Dauer der Arbeitslosigkeit (alo_dau)		67						
Ortsangaben		67						
Wohnort - Kreis (wo_kreis)		67						
Wohnort - Bundesland (wo_bula)		68						
Wohnort - Arbeitsagentur (wo_aa)		68						

Merkmalsliste	BHP Basis	Seite	BeH	LeH	LHG	ASU	XASU	MTH
Wohnort - Regionaldirektion (wo_rd)		69						
Betriebsmerkmale		70						
WZ73 3-Steller (w73_3)	*	70						
WZ93 5-Steller (w93_5)	*	70						
WZ93 3-Steller (w93_3)	*	71						
WZ03 5-Steller (w03_5)	*	71						
WZ03 3-Steller (w03_3)	*	72						
WZ08 5-Steller (w08_5)	*	72						
WZ08 3-Steller (w08_3)	*	73						
W73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen)	*	73						
Art der Vervollständigung w73_3 (group_w73_3)	*	73						
W93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen)	*	74						
Art der Vervollständigung w93_3 (group_w93_3)	*	74						
Jahr des 1. Auftretens der Betriebsnummer (grd_jahr)	*	74						
Erstes Auftreten Betriebsnummer (grd_dat)	*	75						
Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr)	*	75						
Letztes Auftreten Betriebsnummer (lzt_dat)	*	76						
Anzahl Beschäftigte gesamt (az_ges)	*	76						
Anzahl in Vollzeit (Normalbeschäftigte + sonstige) (az_vz)	*	76						
Anzahl geringfügig Beschäftigte (az_gf)	*	77						
Mittelwert imputiertes Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)	*	77						
Arbeitsort Kreis (ao_kreis)	*	78						
Arbeitsort Bundesland (ao_bula)	*	78						

1.5 Mengengerüst

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

Tabelle 5 Mengengerüst

Fallzahlen	vor Episodensplitting	nach Episodensplitting
BeH	344.741	396.736
LeH	35.641	55.889
LHG	9.965	30.257
ASU	52.641	94.632
XASU	2.286	4.872
MTH	9.700	17.447
Sätze insgesamt	454.974	599.833
Personen		12.660

2 Datenquellen

2.1 NEPS-SC6-Erhebungsdaten (NEPS-SC6)

Die NEPS-SC6-Teilstudie „Bildung im Erwachsenenalter und lebenslanges Lernen“ stellt eine Datengrundlage bereit, die Forschung zu Bildungs- und Erwerbsverläufen sowie zur Kompetenzentwicklung im Lebensverlauf von Erwachsenen im erwerbsfähigen Alter, d. h. von 23 Jahren bis zur Verrentung und darüber hinaus ermöglicht. Ziel dieser Teilstudie ist es, Bildung im Erwachsenenalter zu untersuchen und die vorhandenen Wissenslücken über Erwachsenenbildung sowie über Kompetenzausstattung und -veränderung nach der Erstausbildung zu schließen. Dazu wird zum einen die ganze Bandbreite von Bildungsaktivitäten und Lernprozessen (formale, nicht-formale und informelle Bildung), Entscheidungen, die zu ihrer Beteiligung führen, die Aufgaben am Arbeitsplatz und der bisherige Lebensverlauf der Befragten detailliert erfasst. Darüber hinaus werden auch Lese-, Mathematik-, naturwissenschaftliche und ICT-Kompetenzen sowie nicht-kognitive Fähigkeiten wie Persönlichkeit, Motivation und soziale Kompetenzen erhoben.

Inhaltlich sollen die erhobenen Daten es unter anderem ermöglichen,

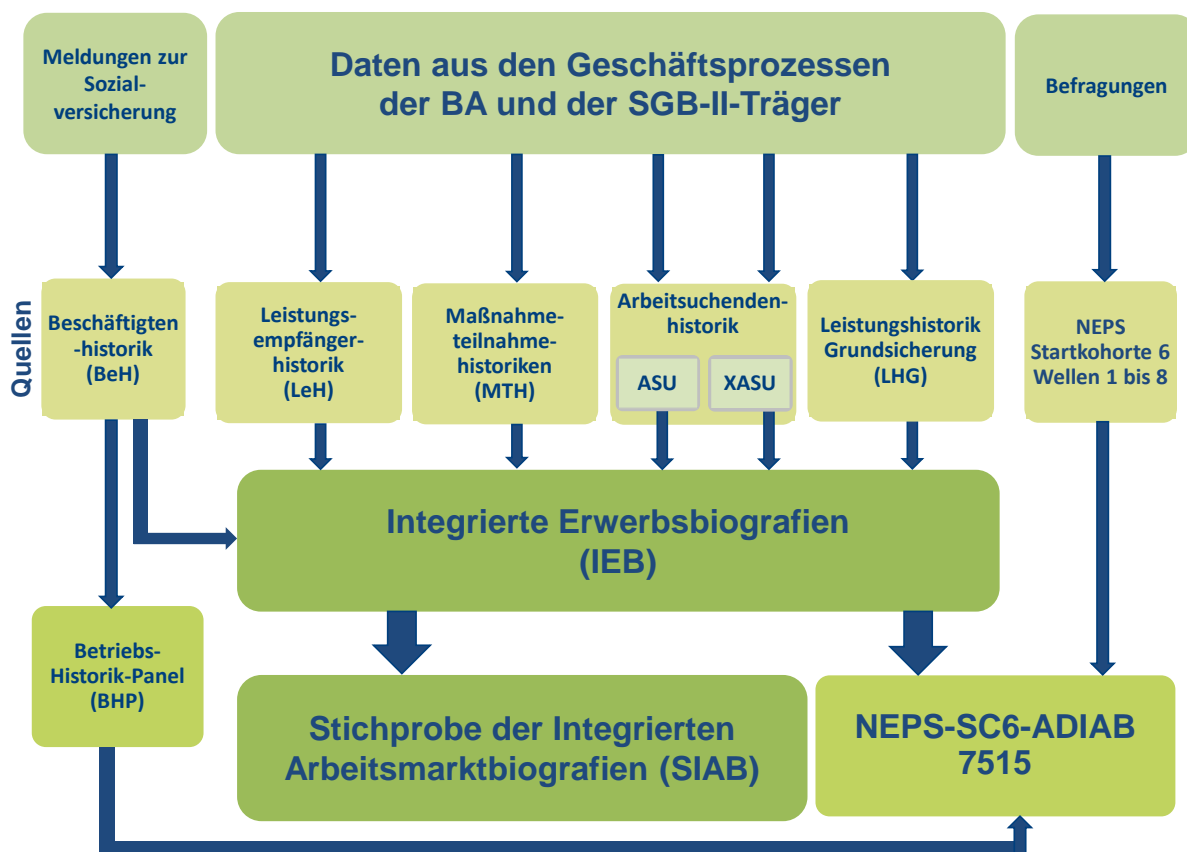
- die Bildungsverläufe von Erwachsenen über ihre Erwerbsgeschichte hinweg und darüber hinaus ins Rentenalter nachzuzeichnen und die Bildungs- und Erwerbsverläufe jüngerer Kohorten nach dem Eintritt ins Erwerbsleben weiterzuverfolgen;
- die Determinanten für Entscheidungen, an formalen oder nicht-formalen Lernaktivitäten nach der Erstausbildung teilzunehmen oder nicht teilzunehmen, zu identifizieren;

- die Kompetenzausstattung unterschiedlicher Gruppen von Erwachsenen in Deutschland zu beschreiben und die Kompetenzentwicklung im Erwachsenenalter sowie die Rolle, die die Erwerbssituation dabei spielt, zu erklären;
- die Wirkungen spezifischer Bildungskontexte im Erwachsenenalter, vor allem der Erwerbssituation und der Familienkonstellation, auf Bildungsentscheidungen und die Weiterbildungsbeteiligung zu analysieren;
- die Erträge von formalen Qualifikationen, Kompetenzen und Erwerbserfahrung in Form von Löhnen, Erwerbskarrieren und Elementen anderer Lebensbereiche wie z. B. dem Wohlbefinden oder dem sozialen Engagement abzuschätzen;
- empirische Ergebnisse zur Kompetenzausstattung von Personen mit Migrationshintergrund, ihren Ressourcen, ihrer Weiterbildungsbeteiligung und deren Erträgen zu generieren;
- Gelegenheiten und Hindernisse für Lernprozesse und Bildung im höheren Erwachsenenalter zu identifizieren.

2.2 Administrative Daten aus den Integrierten Erwerbsbiografien (IEB)

Die administrativen Personendaten werden aus den Integrierten Erwerbsbiografien (IEB) des IAB gezogen und hinsichtlich Variablenauswahl und -ausprägungen entsprechend der vom FDZ erstellten Stichprobe der integrierten Arbeitsmarktbiografien (SIAB, siehe Antoni et al., 2016) aufbereitet. Die IEB vereinen Daten aus fünf verschiedenen Datenquellen, die jeweils Informationen aus unterschiedlichen EDV-Verfahren enthalten können. Zusätzlich werden einige ergänzende Merkmale aus diesen Datenquellen in die administrativen Personendaten aufgenommen, die nicht Teil der IEB sind. Abbildung 2 veranschaulicht die Datenflüsse, die zu NEPS-SC6-ADIAB führen und stellt deren Beziehung zu anderen Datenprodukten des FDZ dar.

Abbildung 2 Quellen von NEPS-SC6-ADIAB 7515



2.3 Beschäftigtenhistorik (BeH)

Die Quelle für Informationen über Beschäftigung ist die Beschäftigtenhistorik (BeH) des IAB. Die Grundlage der Daten ist das mit Wirkung vom 1. Januar 1973 eingeführte (und seit dem 1. Januar 1991 auf Ostdeutschland ausgedehnte) integrierte Meldeverfahren zur Kranken-, Renten- und Arbeitslosenversicherung, das durch das Kürzel DEÜV (früher DEVO/DÜVO) bezeichnet ist (vgl. für weitere Details: Bender et al., 1996, S. 4ff.; Wermter & Cramer, 1988). Es verlangt von den Arbeitgebern, mindestens einmal jährlich Meldungen über alle sozialversicherungspflichtig beschäftigten Arbeitnehmerinnen und Arbeitnehmer an die zuständigen Sozialversicherungsträger abzugeben. Die BeH erfasst alle Arbeiterinnen und Arbeiter, Angestellten sowie Auszubildenden, soweit sie nicht von der Sozialversicherungspflicht befreit sind. Dies bedeutet, dass verbeamtete und selbstständige Tätigkeiten sowie Studierende⁸ (vgl. Cramer, 1985) grundsätzlich nicht in der BeH erfasst werden. Seit der Änderung des Meldeverfahrens zum 1. Januar 1999 werden auch geringfügig Beschäftigte und mithelfende Familien-

⁸ Studierende können dennoch in der BeH erscheinen, wenn sie parallel zum Studium z.B. geringfügig beschäftigt sind.

angehörige erfasst (enthalten erst ab 1. April 1999). Die Daten werden von den Krankenkassen aufgenommen, von der Bundesagentur für Arbeit gesammelt und aufbereitet und anschließend vom IAB in dessen Historikdatei integriert.

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

Bei der Verknüpfung von Personen- mit Betriebsdaten muss berücksichtigt werden, dass die Merkmale in der Basisbetriebsdatei sowie den BHP-Erweiterungsmodulen zum 30.6. eines Jahres aggregiert werden. Die Betriebsmerkmale sind deshalb, im Unterschied zu den Personenangaben, keine zeitraumbezogenen Daten, sondern nur am 30.6. jedes Jahres exakt gültig.⁹

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). Nicht alle Merkmale des BHP sind standardmäßig in NEPS-SC6-ADIAB enthalten. Auf begründeten Antrag hin können einzelne weitere Merkmale des BHP¹⁰ vom FDZ zugespielt werden.

```
use NEPS-SC6-ADIAB_7515_v1.dta, clear
gen int jahr = year(begepi)
sort betnr jahr
merge m:1 betnr jahr using NEPS-SC6-ADIAB_7515_v1_bhp_basis_v1.dta
```

Kasten 1 Beispielcode für Stata

2.4 Leistungsempfängerhistorik (LeH)

Die Leistungsempfängerhistorik (LeH) erfasst Zeiträume, in denen Personen Lohnersatzleistungen der Bundesagentur für Arbeit beziehen (Rechtskreis SGB III). Die Leistungen beinhalten Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld, aber keine Leistungen nach dem Rechtskreis SGB II (z.B. Arbeitslosengeld II). Da der Leistungsanspruch von der Erfüllung gesetzlicher Voraussetzungen abhängt, werden Arbeitslosigkeitsperioden, in denen die Voraussetzungen nicht vorliegen (z.B. fehlende Bedürftigkeit im Falle der Arbeitslosenhilfe oder

⁹ Extrembeispiel: Es liegen Beschäftigungsmeldungen vom 1.1. bis zum 30.5.2006 vor, der Betrieb meldet jedoch Konkurs an, weshalb alle Beschäftigten entlassen werden. Dann gibt es im BHP zu diesem Betrieb für 2006 keine Informationen.

¹⁰ Nähere Informationen zum BHP finden sich bei Schmucker et al. (2016) und unter <http://fdz.iab.de/>.

Nichterfüllung der Anwartschaftszeit beim Arbeitslosengeld) nicht in der Leistungsempfängerhistorik berichtet. Der linke Datenrand der LeH ist der 1. Januar 1975.

2.5 Leistungshistorik Grundsicherung (LHG)

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

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

- Arbeitsgemeinschaften (ARGEn, bis 2010)/ gemeinsame Einrichtungen (ab 2011) aus gemeinsamer Aufgabenwahrnehmung von BA und Kommune,
- Getrennte Trägerschaft/ getrennte Aufgabenwahrnehmung (bis 2011) – hier besteht eine Aufgabenteilung zwischen BA und Kommune¹¹,
- Zugelassene kommunale Träger, die nach der anfänglichen Experimentierklausel des § 6a SGB II auch als optierende Kommunen oder Optionskommunen bezeichnet werden – hier übernimmt die Kommune alle Aufgaben aus dem Bereich des SGB II.

Die Daten der LHG stammen aus verschiedenen Meldeverfahren. In der Regel findet das IT-Verfahren A2LL bis 2010 bei ARGEn und ab 2011 bei gemeinsamen Einrichtungen Verwendung¹². Zugelassene kommunale Träger verwenden verschiedene eigene IT-Verfahren und übermitteln ihre Daten mittels des Standards XSozial-BA-SGB II an die BA. Bei Kommunen mit getrennter Trägerschaft kommen beide Verfahren zum Einsatz. Die unterschiedlichen Datenstandards haben Auswirkungen auf Umfang und Qualität der gelieferten Daten.

¹¹ Die Kommune übernimmt die Kosten der Unterkunft und Heizung (§ 22 SGB II) sowie die abweichend zu erbringenden Leistungen (§ 23 (3) SGB II) und die flankierenden Leistungen nach §16 (2) Satz 2 Nr. 1 - 4 SGB II. Von der BA werden dagegen Regelleistungen, Sozialversicherungsbeiträge und Eingliederungsleistungen (SGB III und SGB II) sowie spezifische Leistungen ohne die oben genannten flankierenden Leistungen erbracht.

¹² 2014 wurde A2LL sukzessive durch ALLEGRO als neues IT-Verfahren für die Grundsicherung im SGB II in gemeinsamen Einrichtungen ersetzt.

Der linke Datenrand der LHG ist der 1. Januar 2005. Die Datenquelle ist jedoch bis Anfang 2007 lückenhaft (siehe Kapitel 5.5).

2.6 Arbeitssuchendehistorik (ASU und XASU)

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

2.7 Maßnahmeteilnahmehistoriken (MTH)

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

3 Datenaufbereitung und Stichprobenziehung

3.1 Bereinigungen und Aufbereitungen

3.1.1 NEPS-SC6-Erhebungsdaten

Detaillierte Informationen zur Stichprobenziehung¹³, zur Datenerhebung, zu den eingesetzten Instrumenten sowie zur Struktur und Aufbereitung der NEPS-SC6-Erhebungsdaten finden sich auf der NEPS-Website. Neben einer Reihe von Reports und Tools für den Umgang mit den Daten steht dort auch das sogenannte „Semantische Datenstruktur-File“ zur Verfügung, welches einen ersten Einblick in die Datensätze der SC6 inklusive Variablennamen, Variablen- und Wertelabels (jedoch ohne Beobachtungen) ermöglicht.¹⁴

¹³ Zur Stichprobenziehung des NEPS siehe Hammon et al. (2016).

¹⁴ <https://www.neps-data.de>

3.1.2 Administrative Daten aus den IEB

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

- Sätze, bei denen das Alter an ihrem Ende kleiner 13 oder zu Beginn größer 75 ist, werden gelöscht.
 - Sätze, bei denen das Enddatum vor dem Beginndatum liegt, werden gelöscht.
 - Widersprüchliche Angaben zu Geschlecht oder Geburtsdatum innerhalb eines Kontos werden bereinigt.
 - Sätze ohne Angabe von Geburtsdatum und Geschlecht nach Bereinigung werden gelöscht.
- Weitergehende Bereinigungen, wie sie z.B. in der IAB-Beschäftigtenstichprobe vorgenommen wurden (Ergänzen von vermutlich fehlenden Meldungen, Streikbereinigungen, etc.), erfolgen nicht.

3.1.3 Beschäftigtenhistorik (BeH)

Um einen im Zeitablauf möglichst konstanten Personenkreis zu erfassen, werden einige Personengruppen, deren Daten nicht für den gesamten Beobachtungszeitraum zur Verfügung stehen, ausgeschlossen. Ab dem Meldejahr 2011 stammen die Daten der BeH aus neu konzipierten Quelldaten. Dadurch wurde eine Reihe von Personengruppen reaktiviert bzw. neu eingeführt, da sie von der BA-Statistik als sozialversicherungspflichtig eingestuft werden. Damit sind fortan die Personengruppen 101 – 107, 111 – 114, 118, 119, 120, 140, 141, 142, 143, 149, 201 und 203 – 205 enthalten sowie die beiden Gruppen 109 und 209, die geringfügig Beschäftigte kennzeichnen. Nicht enthalten sind unter anderem kurzfristig Beschäftigte, d. h. Personengruppen 110, 202 und 210.

- Die Personengruppen 123, 124 und 127 wurden neu eingeführt.
- Aus datenschutzrechtlichen Gründen werden die Personengruppen 107, 111, 113, 114, 127 und 204 zur Personengruppe „Sonstige Erwerbstätige“ (599) zusammengefasst.
- Ab dem Meldejahr 2012 wurden Auszubildende mit den Personengruppen 121 und 122 neu aufgenommen.
- Sätze mit einem Entgelt in Höhe von Null bzw. mit fehlendem Entgelt sowie der Ausprägung 101 beim Merkmal Erwerbsstatus und dem Wert 50 beim Abgabegrund (Jahresmeldung) werden nicht in die IEB übernommen.
- Geschlecht und Geburtsdatum werden aus dem Data Warehouse (DWH) der Statistik der BA übernommen. Dort liegen diese Angaben quellenübergreifend harmonisiert vor.
- Der Gebietsstand für Arbeitsort und Wohnort entspricht dem 31.12.2014.

3.1.4 Leistungsempfängerhistorik (LeH)

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

3.1.5 Leistungshistorik Grundsicherung (LHG)

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

3.1.6 Arbeitssuchendenhistorik (ASU und XASU)

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

3.1.7 Maßnahmeteilnehmehistoriken (MTH)

- Sätze mit einem Enddatum, das vor dem Beginndatum liegt, wurden ausgeschlossen.
- Sätze, die mehr als ein Jahr nach Maßnahmeende erstellt wurden, werden gelöscht, falls ein anderer Satz existiert, der innerhalb eines Jahres nach Abschluss der Maßnahme erstellt wurde.
- Es wird lediglich der aktuellste Satz zu einer einzelnen Maßnahmeteilnahme verwendet.
- In die IEB werden nur Maßnahmeteilnahmen aufgenommen, die als „tatsächlich stattgefunden“ klassifiziert sind. Teilnahmen, die nicht oder noch nicht stattgefunden haben, werden gelöscht. Teilnahmen werden auch dann als nicht stattgefunden gewertet, wenn während der Teilnahme an einer Maßnahme ein Löschdatum gesetzt wird.
- Bestimmte Maßnahmearten werden nicht aufgenommen. Hierzu zählen Unterstützung der Beratung und Vermittlung, Mobilitätshilfen und reine Reha-Maßnahmen.

3.1.8 SGB-II-Anonymisierung

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

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

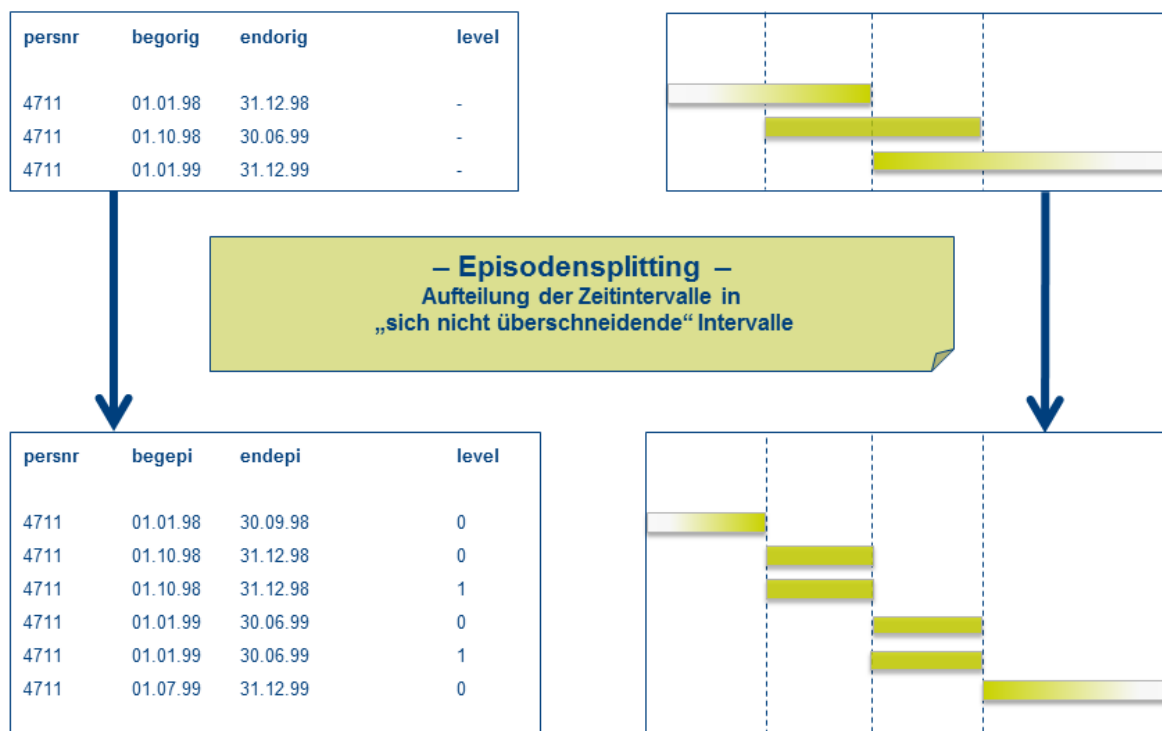
3.2 Episodensplitting

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

Die ursprünglichen Datumsvariablen 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 endepe) festgestellt werden. Um die Originaldaten ohne Episodensplitting wiederherzustellen bzw. die durch das Episodensplitting künstlich erzeugten Datensätze zu löschen, selektiert man alle Sätze, bei denen der Beginn des Originalsatzes gleich dem Beginn der gesplitteten Episode ist (begepi == begorig).

Abbildung 3 Episodensplitting



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

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

Kasten 2 Stata-Code zur Erzeugung von Satzzählern pro Episode

3.3 Fehlende Werte

In den Daten gilt die in Tabelle 6 beschriebene Codierung von fehlenden Werten.

Tabelle 6 Codierung von fehlenden Werten

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

4 Datenverknüpfung

4.1 Informierte Zustimmung der Befragten zur Datenverknüpfung

Ausgangspunkt der Verknüpfung der Erhebungs- und Prozessdaten sind die Befragten der NEPS-SC6. Diese wurden im Rahmen des Interviews um Erlaubnis gebeten, ihre Erhebungsdaten für Forschungszwecke mit den über sie am IAB vorhandenen administrativen Daten verknüpfen zu dürfen. Der Fragebogenschnitt mit der Zustimmungserklärung lautet:

„Um das Interview im Folgenden möglichst kurz zu halten, würden wir gerne bei der Auswertung der Befragung Auszüge aus Daten einbeziehen, die beim Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit in Nürnberg vorliegen. Dabei handelt es sich zum Beispiel um Informationen zu Beschäftigungsverhältnissen, Phasen der Arbeitslosigkeit und Teilnahmen an Maßnahmen während einer Arbeitslosigkeit sowie um betriebliche Merkmale.

Zum Zweck der Zuspiegelung dieser Daten an die Interviewdaten möchte ich Sie herzlich um Ihr Einverständnis bitten.

Werden diese Informationen ausgewertet, so ist absolut sichergestellt, dass alle datenschutzrechtlichen Bestimmungen strengstens eingehalten werden und keine Weitergabe der Daten an Dritte erfolgen wird.

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

Die Befragten mussten der obenstehenden Frage zustimmen, um Teil der Stichprobe für die Verknüpfung zu werden. Die Frage zur Einverständnis zur Verknüpfung wurde in den Wellen 1 bis einschließlich 6 gestellt, wobei Personen diese Frage nur ein weiteres Mal erhielten, wenn sie der Verknüpfung zuvor nicht zugestimmt hatten.

4.2 Herkunft der Verknüpfungsmerkmale

Für die Verknüpfung der Erhebungsdaten der in NEPS-SC6 befragten und zuspielbereiten Personen mit deren administrativen IAB-Daten wurden folgende personenbezogenen Merkmale aus beiden Datenquellen herangezogen:

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

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

Zur Identifizierung der befragten Personen in den administrativen Daten des IAB wurde ein Auszug aus den zentralen Adressdateien des Data Warehouse der Bundesagentur für Arbeit verwendet. Dabei wurde vorab eine Beschränkung auf die in NEPS-SC6 befragten Geburtsjahrgänge und von den Sampling Points abgeleitete Postleitzahl-Bereiche vorgenommen.

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

4.3 Bereinigung und Standardisierung personenbezogener Merkmale

Die oben genannten personenidentifizierenden Merkmale wurden Bereinigungsroutinen unterzogen, um eine möglichst hohe Übereinstimmung zwischen beiden Datenquellen zu erreichen. Für dieses sog. Preprocessing wurden etablierte Verfahren eingesetzt, die am Standort des

German Record Linkage Center (GRLC)¹⁵ am IAB-FDZ für die Besonderheiten des deutschen Kontexts verfeinert wurden. Siehe Christen (2012) für einen umfassenden Überblick über alle in Kapitel 4.3 bis 4.5 beschriebenen Methoden und z. B. Schild & Antoni (2014) für weitere Details zu deren Anpassung an den deutschen Kontext.

Konkret wurden Umlaute und Sonderzeichen durch ihre jeweiligen ASCII-konformen Äquivalente ersetzt, alle Buchstaben in Großbuchstaben umgewandelt sowie überflüssige Leerzeichen und Piktationszeichen entfernt. Mehrere Vor- bzw. Nachnamen wurden jeweils in einem Feld ohne Leerzeichen zusammengefasst. Neben diesen grundlegenden Bereinigungen wurden variablenspezifische Routinen zur Standardisierung der Daten durchgeführt. So wurde die Hausnummer aus einem kombinierten Straße-/Hausnummernfeld extrahiert und alle übrigen Adresszusätze (z. B. Stockwerk, Appartement-Nr.) gelöscht. Bei Straßennamen wurden häufig vorkommende Muster und Abkürzungen standardisiert. Bei Ortsnamen wurden Anhänge auf eine einheitliche Form gebracht (z. B. bei Ortschaften mit zusätzlicher Information wie Fluss oder Region) bzw. entfernt (z. B. bei Berlin-Kreuzberg).

4.4 Exakter Abgleich der Verknüpfungsmerkmale

Nach Bereinigung und Standardisierung der Namens- und Adresskomponenten erfolgte ein exakter Abgleich der Felder zwischen den Einträgen der beiden Datenquellen. Zunächst wurden alle Datenzeilen auf eine exakte Übereinstimmung aller zur Verfügung stehenden Merkmale (Vor- und Nachname, Geburtsdatum, Geschlecht, Straße, Hausnummer, PLZ, Ort) hin geprüft.¹⁶ Im nächsten Schritt wurde untersucht, wie viele weitere Personen sich verknüpfen lassen, wenn eine Abweichung auf einzelnen Adressfeldern oder dem Geburtsdatum toleriert wird (während die Vor- und Nachnamensfelder weiterhin exakt übereinstimmen müssen). Die Toleranz wurde dabei so gewählt, dass ein sehr geringes Risiko bestand, durch dieses Verfahren Einträge zu zwei unterschiedlichen Personen zu verknüpfen (sog. falsch-positive Matches). Es wurden daher zunächst Abweichungen bei einem der besonders fehleranfälligen Merkmale Hausnummer, Postleitzahl, Straße, Ort, Geburtsjahr oder Geburtstag/-monat zugelassen. Als letzten Schritt im Rahmen des exakten Matchings wurden Personen verknüpft, bei denen Vor- und Nachname, Geschlecht, Geburtsdatum und die ersten drei Ziffern der Postleitzahl übereinstimmten, bei denen es aber eine Abweichung auf den Feldern Straße, Hausnummer, Ort und den letzten beiden Ziffern der Postleitzahl gab. Auf diese Weise sollten Personen identifiziert werden, die innerhalb eines bestimmten Umkreises umgezogen waren (und

¹⁵ Siehe Antoni/Schnell (2017) oder <http://www.record-linkage.de> für mehr Information zum GRLC.

¹⁶ Eine Übereinstimmung liegt bezüglich der bereinigten und standardisierten Namens- und Adressfelder vor, nicht bezüglich der unbereinigten Merkmale. Es werden im Rahmen des Preprocessings keine Bereinigungen vorgenommen, die einen Informationsverlust in der Weise nach sich ziehen würden, dass ungleiche Personen einander zugeordnet werden könnten.

dieser Umzug nur in einer der beiden Datenquellen registriert wurde). Die Einschränkung auf dreistellige Postleitzahlgebiete wurde vor dem Hintergrund getroffen, dass auf diese Weise das Risiko von fälschlichen Zuordnungen geringgehalten wurde, da die Wahrscheinlichkeit einer perfekten Übereinstimmung von vollständigem Namen und Geburtsdatum innerhalb solcher regionalen Grenzen sehr gering ist.

4.5 Fehlertoleranter Abgleich der Verknüpfungsmerkmale

Für einen Teil der Personen konnte durch keinen dieser exakten Abgleiche eine Entsprechung in den administrativen Adressdaten ermittelt werden. Um diese Personen mit den administrativen Daten zu verknüpfen, wurden fehlertolerante Linkage-Verfahren verwendet, welche Ähnlichkeitsmaße der Namens- und Adresskomponenten berechnen und die Zeilenpaare mit der höchsten Qualität auswählen. Dabei wurde für jedes abzugleichende Variablenpaar ein Ähnlichkeitsmaß berechnet und diese in einem zusammenfassenden Qualitätsindex aufsummiert. Die Berechnungen wurden mit der Software Merge Tool Box (siehe Schnell et al., 2004) durchgeführt. Für Vor- und Nachname sowie Ort wurde jeweils mittels der Jaro-Winkler-Metrik eine Stringähnlichkeit berechnet, für Straße wird der N-Gram-Algorithmus verwendet¹⁷. Bezüglich der Felder Hausnummer, Postleitzahl, Geburtsdatum (zerlegt in die Komponenten Tag, Monat und Jahr) und Geschlecht wurde ein exakter Vergleich durchgeführt, da Stringähnlichkeitsmaße hier nicht sinnvoll sind.

Beim fehlertoleranten Linkage wurde ein Blocking der Beobachtungen auf Ebene der dreistelligen Postleitzahlgebiete durchgeführt. Dies hat den Vorteil, dass die Anzahl der erforderlichen Vergleiche nicht dem kartesischen Produkt der in den beiden Adressbeständen vorhandenen Zeilen entspricht, sondern auf die Beobachtungen mit identischer dreistelliger PLZ beschränkt bleibt. Zusätzlich wird hierdurch das Risiko der Verknüpfung unterschiedlicher Personen mit sehr ähnlichen Identifikationsmerkmalen verringert. Auf der anderen Seite kann es jedoch in geringem Umfang vorkommen, dass für Personen dadurch keine Entsprechung in den administrativen Daten gefunden werden kann – etwa durch fehlende bzw. fehlerhafte Angaben zur Postleitzahl oder weil ein Umzug in ein anderes Postleitzahlgebiet nur in einer der beiden Datenquellen registriert ist. Zum Vergleich wurde ein zusätzliches Linkage ohne Blocking auf dreistellige Postleitzahlen durchgeführt. Bei den Ergebnissen wurden keine nennenswerten Abweichungen festgestellt. Als Grund hierfür kann gesehen werden, dass Angaben zur Post-

¹⁷ Beim Algorithmus nach Jaro-Winkler werden Ähnlichkeiten am Wortanfang vergleichsweise höher gewichtet, bei N-Grammen fließen hingegen alle Bigramme gleichermaßen in die berechnete Stringähnlichkeit ein. Während Namens- und Ortsfelder oft Zusätze enthalten, die eine höhere Gewichtung des Wortanfangs sinnvoll erscheinen lassen, ist im Falle von Straßennamen von einer vergleichsweise besseren Eignung von N-Grammen auszugehen.

leitzahl in der Praxis selten fehlerhaft sind, vor allem nicht auf den ersten drei Stellen. Außerdem waren Adressänderungen sowohl in den administrativen Adressbeständen als auch in den durch das Befragungsinstitut infas zugeliferten Befragtenadressen sehr gut abgebildet.

Bei der Berechnung des Ähnlichkeitsmaßes wurde ein probabilistisches Linkage angewandt, das rechnerisch berücksichtigt, dass die Wahrscheinlichkeit einer Übereinstimmung von Merkmalen von deren Varianz in der Grundgesamtheit abhängig ist. Hierbei werden so genannte m- und u-Wahrscheinlichkeiten berechnet, die angeben, mit welcher Wahrscheinlichkeit die Angaben zu einer Person zwischen zwei verschiedenen Quellen variieren bzw. wie wahrscheinlich das Auftreten eines identischen Merkmals bei unterschiedlichen Personen ist. Beispielsweise deutet eine Übereinstimmung bei der Variable Geschlecht weniger auf einen tatsächlichen Match hin als eine Übereinstimmung beim Nachnamen und wird deshalb niedriger gewichtet. Die variablenspezifischen m- und u-Gewichte, die dem probabilistischen Linkage zugrunde lagen, wurden im Rahmen vergangener Projekte des GRLC ermittelt. Es wurden alle zur Verfügung stehenden Merkmale zwischen den beiden Datenquellen verglichen und ein additives Ähnlichkeitsmaß berechnet. Je höher der Wert dieses Index, desto größer die Gesamtähnlichkeit der beiden Zeilen und desto wahrscheinlicher liegt eine tatsächliche Übereinstimmung vor.

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

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

Geschlecht der Person. Eine manuelle Sichtung ergab jedoch, dass es sich in diesen Fällen um plausible Matches handelte, bei denen im Ursprungsmaterial Fehlangaben vorlagen.

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

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

4.6 Ergebnis der Datenverknüpfung

Insgesamt 15.982 Personen erteilten bis einschließlich Welle 5 ihre Zustimmung zur Verknüpfung (93,24% aller Befragten). Davon konnten jedoch nur 14.065 Personen für die Datenverknüpfung verwendet werden, da Personen, die ihre Panelbereitschaft zwischenzeitlich zurückgezogen haben, für das Linkage nicht zur Verfügung standen, obgleich sie dieser zuvor zugestimmt hatten.

Für 12.660 Personen konnten die Erhebungsdaten einem Konto in den administrativen Daten zugeordnet werden. Dies entspricht einer sehr hohen Verknüpfungsquote von 90,1% der Personen, die für eine Zuspiegelung der administrativen Daten zur Verfügung standen. Es verblieben lediglich 1.405 Personen (9,99%), die trotz Verknüpfungserlaubnis nicht in den administrativen Datenbeständen gefunden wurden. In Tabelle 7 werden die Erfolgsquoten der einzelnen Verknüpfungsschritte zusammengefasst.

Generierte technische Merkmale im Datensatz (siehe Kapitel 6.2) kennzeichnen für jede befragte Person, mit welcher Verknüpfungsmethode und mit welchem Maß an Sicherheit eine erfolgreiche Verknüpfung zustande gekommen ist. Damit kann jede Nutzerin und jeder Nutzer nach eigenem Ermessen entscheiden, welche der Fälle in den eigenen Analysen Verwendung finden sollen.

Tabelle 7 Ergebnisse der Linkage-Schritte

	N	Anteil an Zustimmenden	Anteil an Befragten
Befragte	17.140		100,00%
Befragte mit Verknüpfungserlaubnis	15.982	100,00%	93,24%
darunter verfügbare Befragte	14.065	100,00%	82,06%
Matches mit IEB-Sätzen	12.660	90,01%	73,86%
darunter:			
Exaktes Matching	11.632	82,70%	67,86%
Probabilistisches Matching	1.012	7,20%	5,90%
Manuelles Matching	16	0,11%	0,09%

5 Datenqualität und Probleme

5.1 NEPS-SC6-Befragungsdaten (NEPS-SC6)

Eine präzise Beschreibung der Daten, deren Struktur sowie Informationen zur Stichprobe und zur Befragung allgemein finden sich im Data Manual der Startkohorte 6. Weiterhin sind dort genierte Variablen und Details zur Gewichtung aufgeführt¹⁸.

5.2 Integrierte Erwerbsbiografien (IEB)

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

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

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

Diese Lücken wurden anhand der Ausprägungen der Merkmale Abgabegrund/Beendigungsgrund/SGB-II-Einstellungsgrund/Abmeldegrund und Erwerbsstatus in den verschiedenen Quellen identifiziert (s. Tabelle 8). Die Auflistung erhebt keinen Anspruch auf Vollständigkeit.

¹⁸ <https://www.neps-data.de>

Tabelle 8 Biographische Lücken und Identifikationsmöglichkeiten

Biografische Lücke	Information über Lücke, ggf. identifizierbar anhand der Ausprägung des Merkmals „grund“ im Vorgängersatz der Quelle
Beamte, Berufssoldaten, Richter, Angestellte von Körperschaften oder Stiftungen des öffentlichen Rechts	XASU
Ungeförderte Selbstständige	LeH, ASU
Studierende, Personen in schulischer Weiterbildung	LeH, LHG, ASU, XASU
Personen, die länger als 6 Wochen krank/ arbeitsunfähig sind (Krankheit während Arbeitslosigkeit ist aber unter bestimmten Bedingungen in der Quelle ASU abgebildet, s. Abschnitt 3.1.6)	BeH, LeH, ASU
Personen in Altersrente ohne Beschäftigung wenn nicht Mitglied in einer BG	LeH, LHG, ASU
Personen in Mutterschutz/ Elternzeit	XASU
Bezieher von Vorruhestandsgeld	LeH, ASU
Hausgewerbetreibende	
Kurzfristig Beschäftigte	ASU
Personen in Einrichtungen der Jugendhilfe, Berufsbildungswerken, anerkannten Werkstätten oder ähnlichen Einrichtungen für behinderte Menschen	ASU
Teilnehmende an Leistungen zur Teilhabe am Arbeitsleben (Rehabilitanden)	ASU
(Nebenerwerbs-)Landwirte	
Pflegepersonen i.S. von § 19 SGB XI	
Grundwehrdienstleistende	BeH, LeH, LHG, ASU, XASU
Wehrübungsleistende	BeH, LeH, LHG, ASU, XASU
Zivildienstleistende	BeH, LeH, LHG, ASU, XASU
Ableistende eines freiwilligen sozialen bzw. ökologischen Jahres anstelle des Zivildienstes	
Sonstige Personen, die nicht bei der Rentenversicherung oder der BA registriert sind (z.B. Sabbatjahr, Finanzierung aus privatem Vermögen oder Rentenbezügen, Emigration, Beschäftigung im Ausland, ehrenamtliche Tätigkeit, etc.)	BeH, LeH, ASU
Streikende bei Streikdauer über einem Monat	LeH
Sozialhilfe-Bezieher (vor Einführung des SGB II 2005), Sozialgeld-Bezieher (im Sinne des SGB II)	
Ausgleichsempfänger nach dem FELEG (Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit)	

5.3 Beschäftigtenhistorik (BeH)

- Die Einführung des neuen Tätigkeitsschlüssels in 2011 hat zu einer Reihe von Problemen geführt. So kam es während des Übergangszeitraums, der Arbeitgebern im Meldeverfahren zur Sozialversicherung eingeräumt wurde,¹⁹ zu einer temporären Zunahme fehlender Angaben. Analysen der Statistik der BA (Bertat et al., 2013) zeigen, dass die neuen bzw. umgestellten Merkmale „ausgeübte Tätigkeit“, „Arbeitszeit“ und „Ausbildungsabschluss“ nach der Umstellung zu etwa 20 bis 30 Prozent keine Angaben enthielten. Eine deutliche Verbesserung stellt sich im ersten Halbjahr 2013 ein. Um die Qualität der Arbeitszeitangabe im Übergangszeitraum zu verbessern, haben Ludsteck und Thomsen (2016) ein Imputationsverfahren entwickelt, das die fehlenden durch imputierte Werte ersetzt. Die imputierten Daten werden in NEPS-SC6-ADIAB 7515 bereits verwendet.
- Aufgrund der Einführung des Meldeverfahrens in den neuen Bundesländern kann erst ab 1993 von einer hinreichenden Vollständigkeit der Meldungen ostdeutscher Betriebe ausgegangen werden. Aus demselben Grund weisen sehr viele Sätze des Jahres 1991 für einige Merkmale (u.a. Erwerbsstatus) fehlende Werte auf.
- Der Anstieg der Zahl der BeH-Sätze ab 1999 folgt aus der Einführung der Meldepflicht für geringfügig Beschäftigte ab dem 1. April 1999.
- Es lässt sich insbesondere 1999 ein deutlicher Anstieg bei den Meldungen zu Teilzeitbeschäftigung feststellen. Verursacht wird dieser Befund sowohl vom tatsächlich zu beobachtenden Anstieg der Teilzeitbeschäftigung als auch von der Tatsache, dass seit 1999 die Beschäftigtenmeldungen generell besser ausgefüllt werden.
- Innerhalb des Meldeverfahrens ist eine gewisse Zeitverzögerung nicht vermeidbar. Zwar müssen Änderungen von Beschäftigungsverhältnissen sofort gemeldet und bestehende Beschäftigungsverhältnisse jährlich bis April (bis Jahresmeldung 2012) bzw. Mitte Februar (ab Jahresmeldung 2013) des folgenden Jahres bestätigt werden. Faktisch treffen Meldungen jedoch zum Teil um Jahre verspätet ein. Die Historikdatei des IAB wird aber nicht kontinuierlich fortgeschrieben, sondern in bestimmten Abständen aktualisiert. Dies geschieht unter Verwendung von Dateien mit den Beschäftigten-Meldungen jeweils eines Jahres, die 36, 18, 12 oder 6 Monate nach Ablauf des Berichtsjahres eingetroffen sind (frühestens im Juli 2015 z.B. kann die 18-Monats-Datei für das Jahr 2013 erstellt werden). Mehr als 3 Jahre verspätete Meldungen werden im IAB nicht berücksichtigt, so dass eine 36-Monats-Datei per definitionem einen Füllgrad von 100% aufweist.

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

- In der zugrundeliegenden Version der IEB liegt der Füllgrad der BeH-Meldungen zum Jahr 2012 bei knapp 100%, da für dieses Jahr nur das 30-Monatsmaterial verwendet werden konnte.²⁰ Für das Jahr 2013 wurde eine 18-Monats-Datei verwendet, für 2014 eine 6-Monats-Datei. Es ist daher davon auszugehen, dass die BeH-Meldungen für das Jahr 2013 leicht und für 2014 etwas stärker untererfasst sind. Die Auswertbarkeit der Daten auf Personenebene sollte dadurch aber nicht eingeschränkt sein. Allerdings häufen sich die fehlenden Meldungen in wenigen Betrieben. Dies bedeutet, dass in Einzelfällen die Betriebsangaben, z.B. die Betriebsgröße, grob falsch sind und sich in Folgeversionen stark verändern werden.
- Im Jahr 1984 erfolgte eine Änderung des Meldeverfahrens. Einmalig gezahltes Bruttoarbeitsentgelt wurde ab diesem Zeitpunkt als Teil des beitragspflichtigen Jahresentgelts gemeldet, was zu einer Erhöhung des durchschnittlichen Tagesentgelts führt. Insbesondere steigt ab diesem Jahr der Anteil der Entgelte oberhalb der Beitragsbemessungsgrenze stark an (vgl. Bender et al., 1996).
- Für die Jahre 1992 bis 2000 wurden in den IEB auffällige Einbrüche und Anstiege in der Anzahl der Meldungen festgestellt. Es lassen sich v.a. Einbrüche bei folgenden zehn Kreisen beobachten: Braunschweig (03101), Wolfsburg (03103), Emden (03402), Kassel (06633), Essen (05113), Neuss (05162), Rhein-Erft-Kreis (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 zur 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 (Ärztinnen und Ärzte, Apothekerinnen und Apotheker) in der Variable Beruf – ausgeübte Tätigkeit (beruf) im Vergleich zu benachbarten Jahren sehr selten vor. Die Gründe dafür sind nicht bekannt.

²⁰Durch 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.

5.4 Leistungsempfängerhistorik (LeH)

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

5.5 Leistungshistorik Grundsicherung (LHG)

- In den Jahren 2005 und 2006 ist die LHG nicht vollständig gefüllt. Es kommt zu erheblichen trägerspezifischen Ausfällen. Es wird daher abgeraten, für diesen Zeitraum Analysen allein auf Basis der LHG zu machen.
- Längsschnittuntersuchungen für Personen sind mit Unsicherheiten behaftet, da Umzüge in und aus Gebieten von Trägern mit Lieferproblemen nicht von Änderungen des Status der Leistungsberechtigung unterschieden werden können.
- Auch ab 2007 kommt es hauptsächlich in der LHG immer wieder zu meist einmonatiger trägerspezifischer Untererfassung.
- Im Zusammenhang mit Trägerartwechseln kommt es zu Unter-/Übererfassungen:
 - Im Rahmen der Trägergebietsreform vom 1. Januar 2011 kommt es zu Untererfassungen im Gebiet der Arbeitsagenturen Dessau-Roßlau, Halberstadt, Halle und Sangerhausen.
 - Doppelmeldungen aufgrund der Gebietsreformen 2009/2011 und der Trägerformwechsel zum 1. Januar 2012 werden in der IEB bereits soweit möglich bereinigt. Dennoch können Doppelmeldungen auftreten.
- Bei folgenden Jobcentern bestehen Unschärfen bei der Zuordnung von Leistungsfällen:
 - zwischen Emden und Norden zwischen September und Dezember 2009
 - zwischen Döbeln und Mittelsachsen von Oktober bis Dezember 2012
 - zwischen Tirschenreuth und Wunsiedel von November 2012 bis März 2013
- Manche Personen, für die eine LHG-Beobachtung vorliegt, sind vom Leistungsbezug nach SGB II ganz oder teilweise ausgeschlossen, da sie z. B. eine geförderte Ausbildung machen, Altersrente beziehen, in einer stationären Einrichtung untergebracht sind oder Versicherungsleistungen zur Vermeidung von Hilfebedürftigkeit bekommen. Dies betrifft pro Jahr

ca. 3 bis 5 Prozent der Fälle. In XSozial kommt es zu trägerspezifischen Untererfassungen dieser Personengruppe. In NEPS-SC6-ADIAB ist der Ausschluss von Leistungen nicht identifizierbar.

5.6 Arbeitssuchendehistorik (ASU/XASU)

5.6.1 ASU

- Die registrierten Zeiten der Arbeitssuche der Quelle ASU können ab 1997 als vollständig angesehen werden. Das Analysepotenzial der ASU-Sätze vor 1997 ist damit eingeschränkt.
- Für die Vermittlungskräfte ist es nicht möglich, die Zuordnung zum Rechtskreis immer unmittelbar zu erfassen, da vielfach erst zeitverzögert feststeht, welcher Träger aufgrund einer möglichen SGB-II-Leistungsberechtigung vorrangig zuständig ist. Es wird daher empfohlen, die Ausprägung des Merkmals „Trägerart“ in der ASU mit der Ausprägung in der LHG im gleichen Zeitraum abzugleichen. Aufgrund der Erfassungslücken bei LHG im Zeitraum 2005 bis 2006 ist dies nicht immer möglich.
- Für manche Personen, für die ab 2005 eine Optionskommune zuständig war, wurden von der BA parallel "künstliche" ASU-Sätze angelegt. Diese sind über estatvor (Wechsel zu einem zugelassenen kommunalen Träger) zu identifizieren.
- Von Mitte 2005 bis Mitte 2006 wurde in der BA das EDV-Verfahren coArb, aus dem die Arbeitssuchendaten stammen, durch das Verfahren VerBIS abgelöst. Ab Juli 2005 wurde coArb erstmals in der Agentur Wiesbaden als Pilotprojekt, ab Dezember 2005 dann sukzessive in mehreren Tranchen in allen Arbeitsagenturen durch VerBIS ersetzt. Viele der erfassten Merkmale wurden in den beiden Systemen in unterschiedlicher Differenziertheit und unterschiedlicher qualitativer Gewichtung abgefragt. Die Integration dieser Merkmale in die IEB gestaltet sich daher schwierig und ist nur durch die Anwendung eines speziellen Verfahrens (Mapping) möglich. Leider kann durch das Mapping keine vollständige Umsetzung der betroffenen Merkmale von coArb auf VerBIS erreicht werden, sodass sich bei einigen Merkmalen die Ausprägungen „Keine Angabe“, „Sonstiges“ oder „Missing“ ungewöhnlich häufen. Außerdem können sich bei Häufigkeitsauszählungen auffällige Unterschiede ergeben, je nachdem ob die Ursprungsquelle der Daten coArb oder VerBIS war. Wichtige Einschränkungen des Analysepotentials werden in der jeweiligen Variablenbeschreibung angesprochen.
- Das bis Juni 2006 genutzte Verfahren coArb unterstützte ausschließlich die Vermittlung von Arbeitslosen und Arbeitssuchenden. Lückenhaft erfasst wurden auch Daten zu Personen, die sich lediglich beraten lassen wollten. Die Daten der Berufsberatung wurden in ei-

nem separaten System gehalten. Mit VerBIS erweiterten sich die Ausprägungen des Arbeitsuche-Status um „Ratsuchende“ und Personen „ohne Status“. Zu Letzteren gehören ALG-II-Leistungsberechtigte, die der Arbeitsvermittlung nur eingeschränkt 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.

5.6.2 XASU

- Im Gegensatz zu den Arbeitsuchendendaten aus den ARGEn und den getrennten Trägerschaften zeigen sich nach 1. Januar 2005 für die zugelassenen kommunalen Träger zwischenzeitlich systematische Untererfassungen. Deshalb sollten Daten aus der Quelle XASU erst ab 2007 ausgewertet werden.
- Verschiedene Merkmale weisen für die XASU zum Teil nur einen sehr geringen Füllgrad auf. Hiervon betroffen sind unter anderem die Merkmale Schulabschluss, Abgabegrund sowie Erwerbsstatus vor Arbeitsuche. Die Füllgrade dieser Merkmale verbessern sich zwar im Zeitablauf, bleiben aber dennoch zum Teil unbefriedigend. Das Merkmal Beruf ist nahezu für den gesamten zur Verfügung stehenden Zeitraum in der XASU nicht gefüllt.
- Bei etlichen Trägern (Kreisen) erscheint der Anteil der gemeldeten Arbeitsuchenden an den gemeldeten erwerbsfähigen Hilfebedürftigen in den IEB durchgängig oder zeitweise unplausibel hoch. Gründe dafür könnten eine möglicherweise fehlerhafte Bestimmung des Status „Nichtarbeitslos, arbeitssuchend“ oder eine technisch fehlerhafte Übermittlung mit hohem Anteil an Nachmeldungen bei diesen Trägern sein.
- Die träger- und zeitraumabhängige Plausibilität der XASU-Angaben sollte vor der Verwendung jeweils unter Berücksichtigung der Untersuchungsfrage geprüft werden.

5.7 Maßnahmeteilnahmehistoriken (MTH)

- Die MTH sind für Maßnahmen mit einem Eintrittsdatum vor dem 1. Januar 2000 unvollständig.
- Zum 1. Januar 2005 kommt es zu einem Bruch, da mit Einführung des SGB II Maßnahmeteilnehmerinnen und -teilnehmer auf verschiedene Träger aufgeteilt wurden (siehe Kapitel 2.5 und 2.7).
- Die MTH enthält nur Meldungen, die in BA-Verfahren gebucht sind. Deren Nutzung nimmt in Arbeitsgemeinschaften/getrennten Trägerschaften/getrennten Aufgabenwahrnehmungen zwischen 2005 und 2007 kontinuierlich zu. Ab März 2007 sind die Meldungen für diese Träger komplett.

- Maßnahmen, die über den Standard XSozial gemeldet werden, sind nicht in der MTH oder NEPS-SC6-ADIAB enthalten.
- Durch die Neuorganisation von SGB-II-Trägern kann die Dokumentation von Maßnahmeteilnahmen in der MTH abrechnen oder neu beginnen, wenn es zu einer Umstellung des Meldeverfahrens kommt.²¹
- Bei Meldungen zum Überbrückungsgeld wird zum Teil die maximal zulässige Dauer von einem halben Jahr überschritten. Dies kann in den meisten Fällen durch eine Voreinstellung der Eingabemaske im Erfassungssystem erklärt werden.
- Die MTH wird um Bewerbermerkmale (z. B. Ausbildung) aus anderen BA-Verfahren ergänzt. Für diese Merkmale erfolgte im Jahr 2006 die Umstellung des operativen Verfahrens von coArb auf VerBIS. Es gelten hier die gleichen Qualitätseinschränkungen wie für die ASU.

6 Variablenbeschreibung

Auszählungen sowie Übersichten über die einzelnen Ausprägungen und Labels der Merkmale finden sich in den Arbeitshilfen zu NEPS-SC6-ADIAB unter <http://fdz.iab.de>.

6.1 Identifikatoren

6.1.1 Personen-ID für NEPS-SC6-ADIAB (ID_t)

Variablenlabel	Personen-ID für NEPS-SC6-ADIAB
Variablenname	ID_t
Kategorie	Identifikatoren
Herkunft	Generiert
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Die Personen-ID zeigt an, welche Sätze zu einer Person gehören. Sie zeigt auch an, welche Personen in Erhebungs- und administrativen Daten korrespondieren, d. h. die ID_t ist sowohl in den Erhebungsdaten als auch in den administrativen Daten enthalten. Diese Identifikationsvariable ist systemfrei, weshalb sie keine Rückschlüsse auf Eigenschaften der Person oder Original-Identifikatoren ermöglicht.

6.1.2 Betriebs-ID für NEPS-SC6-ADIAB (betnr)

Variablenlabel	Betriebs-ID für NEPS-SC6-ADIAB
Variablenname	betnr
Kategorie	Identifikatoren

²¹ Weitere Informationen zur Gebietsstruktur der SGB-II-Träger und relevanten Änderungen finden sich unter: <https://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Regionale-Gliederungen/Gebietsstruktur-Traeger-Grundsicherung-Nav.html>

Herkunft	Generiert
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Die Betriebs-ID zeigt an, welche Sätze zum gleichen Betrieb gehören. Sie basiert auf der von der BA vergebenen Betriebsnummer, die durch eine systemfreie Nummer ersetzt wurde. (Weitere Informationen zur Betriebsnummernvergabe der BA finden sich in Bender et al., 1996, S. 15f. und S. 27-30.) Die Betriebsnummer kann zusammen mit der Jahresangabe genutzt werden, um Personen- und Betriebsinformationen zusammenzuspielen.</p> <p>Für die Betriebsnummer sollte Folgendes grundsätzlich beachtet werden:</p> <ul style="list-style-type: none"> a) Besteht das Unternehmen aus nur einer Niederlassung oder hat das Unternehmen in einer Gemeinde nur eine Niederlassung, so ist diese Niederlassung der Betrieb und erhält eine Betriebsnummer. b) Hat das Unternehmen mehrere Niederlassungen in einer Gemeinde, so sind diese Betriebs- und Arbeitsstätten zu einem Betrieb unter einer Betriebsnummer zusammenzufassen, wenn sie derselben Wirtschaftsklasse angehören bzw. ist jede Niederlassung ein Betrieb und erhält je eine Betriebsnummer, wenn sie nicht derselben Wirtschaftsklasse angehören. c) Hat das Unternehmen mehrere Niederlassungen in mehreren Gemeinden, so ist jede dieser Niederlassungen ein Betrieb und erhält je eine Betriebsnummer. <p>In diesem Zusammenhang sind auch folgende Definitionen im Sinne der Betriebsnummernvergabe im Rahmen des Meldeverfahrens zur Sozialversicherung zu beachten:</p> <ul style="list-style-type: none"> a) Ein Betrieb ist eine regional und wirtschaftsfachlich abgegrenzte Einheit, in der Beschäftigte tätig sind und für die nach o.a. Grundsätzen eine Betriebsnummer zu vergeben wird. b) Eine Arbeitsstätte ist eine Einheit, in der Beschäftigte tätig sind und für die nach o.a. Grundsätzen eine Betriebsnummer nicht zu vergeben ist. c) Ein Unternehmen als Begriff fasst Betriebs- und Arbeitsstätten desselben Arbeitgebers zusammen. d) Ein Arbeitgeber ist jede natürliche oder juristische Person, für die mindestens ein(-e) sozialversicherungspflichtig bzw. geringfügig Beschäftigte oder Beschäftigter tätig ist. e) Betrieb und Betriebsstätten sind synonyme Begriffe; Niederlassung ist dann ein synonyme Begriff für Filialen, Zweigstellen, Auslagerungen, Dependancen, Arbeitsstätten u. Ä., wenn sie kein Betrieb ist.

Qualitätshinweis	Die Betriebsnummer fehlt in nur sehr wenigen BeH-Meldungen. Bei diesen Meldungen handelt es sich in den meisten Fällen um Meldungen der Personengruppe „205“ (Entgeltmeldungen für unständig Beschäftigte). Da Betriebsmerkmale (Arbeitsort, Wirtschaftszweig, Betriebsgröße usw.) über die Betriebsnummer zugespielt werden, fehlen sie in diesen Meldungen.
------------------	---

6.2 Generierte technische Merkmale

6.2.1 Satzzähler pro Konto (spell)

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

6.2.2 Quelle des Satzes (quelle)

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

6.2.3 Jahr (jahr)

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

6.3 Gültigkeitszeitraum

6.3.1 Beginndatum Originalbeobachtung (begorig)

Variablenlabel	Beginndatum Originalbeobachtung
Variablenname	begorig
Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	<p>Das Beginndatum des Originalsatzes entspricht dem ursprünglichen Beginndatum der Meldung. Dies kann vom Beginndatum der betreffenden Datenzeile, dem sogenannten Episoden-Beginndatum, abweichen (vgl. auch die Erläuterungen zum Episodensplitting in Kapitel 3.2). Da die Meldelogik unter bestimmten Umständen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben in diesen Fällen durch das in Kapitel 3.1.8 beschriebene Anonymisierungsverfahren verändert.</p> <p>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.</p> <p>2) LHG, ASU, XASU Bestimmte Änderungen führen zur Anlage eines neuen Datensatzes (s. Kapitel 2.5 und 2.6). begorig gibt das Beginndatum des neuen Zeitraums an.</p>

6.3.2 Enddatum Originalbeobachtung (endorig)

Variablenlabel	Enddatum Originalbeobachtung
Variablenname	endorig
Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	<p>Das Enddatum des Originalsatzes entspricht dem ursprünglichen Enddatum der Meldung. Dies kann vom Enddatum der betreffenden Datenzeile, dem sogenannten Episoden-Enddatum, abweichen (vgl. auch die Erläuterungen zum Episodensplitting in Kapitel 3.2). Da die Meldelogik unter bestimmten Umständen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben in diesen Fällen durch das in Kapitel 3.1.8 beschriebene Anonymisierungsverfahren verändert.</p> <p>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.</p>

	2) LHG, ASU, XASU Bestimmte Änderungen führen zur Anlage eines neuen Datensatzes (s. Kapitel 2.5 und 2.6). begorig gibt das Beginndatum des neuen Zeitraums an.
--	---

6.3.3 Beginndatum der Episode (begepi)

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

6.3.4 Enddatum der Episode (endepe)

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

6.4 Informationen zur Person

6.4.1 Geschlecht (frau)

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

6.4.2 Geburtsjahr (gebjahr)

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

Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Geburtsjahr ist innerhalb eines Personenkontos konstant.
Qualitätshinweis	In den Originaldaten kann es vorkommen, dass das Geburtsjahr zwischen den Quellen wechselt. Dies wird bei der Datenaufbereitung bereinigt. Hierbei kommt Angaben aus dem Leistungsbezug die höchste Priorität zu.

6.4.3 Staatsangehörigkeit (nation)

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

6.4.4 Staatsangehörigkeit vergrößert (nation_gr)

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

6.4.5 Familienstand (famst)

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

6.4.6 Kinderzahl (kind)

Variablenlabel	Kinderzahl
Variablenname	kind
Kategorie	Persönliche Merkmale
Herkunft	LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Je nach Datenquelle hat das Merkmal eine unterschiedliche Bedeutung.</p> <p>1) LeH In der LeH gibt die Variable die Anzahl der Kinder unter 16 Jahren zum Zeitpunkt der Antragstellung an. Es wird nur unterschieden, ob die Anzahl an Kindern gleich Null oder mindestens eins ist. Folgende Ausprägungen kommen daher vor: 0 Keine Kinder 100 Ein oder mehrere Kinder Das Merkmal wird bei Änderungen der Leistungsart oder neuen Bewilligungen nicht aktualisiert, sondern erst, wenn ein erneuter Leistungsfall auftritt. Dadurch ist die Datenqualität eingeschränkt.</p> <p>2) ASU, MTH Hier entspricht die Ausprägung des Merkmals der Anzahl an Kindern unter 15 Jahren. Bis zum 30. Juni 2006 konnten nur bis zu neun Kinder erfasst werden. Der Wert 0 kommt nicht vor. Vor dem 30. Juni 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. Juni 2006 ist das Merkmal nur gefüllt, wenn Kinder vorhanden sind.</p> <p>3) XASU, LHG Hier entspricht die Ausprägung des Merkmals der Anzahl an Kindern unter 15 Jahren in der Bedarfsgemeinschaft. Bei der Quelle LHG gilt der Wert für den gesamten Originalzeitraum.</p>

6.4.7 Ausbildung (ausbildung)

Variablenlabel	Ausbildung
Variablenname	ausbildung
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine

Detailbeschreibung	<p>Zu beachten ist, dass das Merkmal je nach Quelle eine unterschiedliche Bedeutung hat.</p> <p>1) BeH</p> <p>Das Merkmal enthält für BeH-Sätze die von den Arbeitgebern im Rahmen des Meldeverfahrens anzugebende Berufsausbildung. Folgende Ausprägungen kommen vor:</p> <ol style="list-style-type: none"> 1 ohne abgeschlossene Berufsausbildung 2 betriebliche/außerbetriebliche Ausbildung 11 Fachhochschulabschluss 12 Hochschulabschluss <p>Bei Meldungen nach dem neuen Tätigkeitsschlüssel ist es nicht mehr möglich, FH-Absolventen wie bisher eindeutig zu identifizieren, da der neue Tätigkeitsschlüssel die separate Erfassung dieses Bildungsabschlusses nicht länger erlaubt. Sie werden der Kategorie 12 zugeordnet. „Zusammen mit einem Betriebswechsel treten häufig auch Änderungen im Ausbildungsstatus auf. Das liegt daran, dass beim neuen Betrieb die Meldedaten neu zusammengestellt werden. Wenn ein Beschäftigter z.B. durch berufsbegleitende Weiterbildung einen höheren Abschluss erlangt hat, dann wird dieser Statuswechsel wahrscheinlich erst von einem neuen Beschäftigungsbetrieb erfasst. Es ist in der Regel davon auszugehen, dass bei längeren Beschäftigungszeiten innerhalb eines Betriebs die einmal erhobenen personenbezogenen Daten fortgeschrieben werden.“ (Meinken/Koch, 2004, S. 63)</p> <p>Eine Methode, fehlende Werte oder inkonsistente zeitliche Verläufe der Bildungsangaben in der IAB-Beschäftigtenstichprobe (IABS) zu bereinigen, findet sich in Fitzenberger et al. (2006), sowie in Drews (2006). Hierfür wird aber nur die Quelle BeH herangezogen, die in der IABS die einzige Datenquelle mit Informationen zur Bildung war.</p> <p>2) ASU, XASU, MTH</p> <p>Für diese Sätze ist die letzte abgeschlossene Berufsausbildung enthalten. Folgende Ausprägungen kommen vor:</p> <ol style="list-style-type: none"> 1 ohne abgeschlossene Berufsausbildung 2 betriebliche/außerbetriebliche Ausbildung 3 Berufsfachschule 4 Fachschule 5 Fachhochschule 6 Universität 7 in Deutschland nicht anerkannte Berufsausbildung 8 in Deutschland nicht anerkannter Hochschulabschluss <p>Wobei die Ausprägungen 7 und 8 nur für die Quelle MTH gelten.</p>
--------------------	--

Qualitätshinweis	<p>Die Zahl der fehlenden Werte steigt im Zeitverlauf kontinuierlich an und erreicht am rechten Datenrand über 40%. Fehlende Werte treten besonders häufig bei folgenden Personengruppen auf: geringfügig beschäftigten Personen, Personen in Teilzeit, ausländischen Beschäftigten und ostdeutschen Arbeitnehmerinnen und Arbeitnehmer. Hintergrund ist, dass das Merkmal versicherungsrechtlich keine besondere Bedeutung hat (s. Meinken & Koch, 2004, S. 63).</p> <p>Durch die Umstellung von coArb auf VerBIS kann in den Quellen ASU und MTH zwischen 2006 und 2008 nicht korrekt zwischen „ohne abgeschlossene Berufsausbildung“ und „Keine Angabe“ unterschieden werden. Ein fehlender Wert in diesem Zeitraum besagt also nicht unbedingt, dass keine (Information zur) Ausbildung vorliegt, sondern kann auch bedeuten, dass das einschlägige Generierungsverfahren nicht angewandt werden konnte. Der Füllgrad in der Quelle XASU ist generell niedrig.</p>
------------------	--

6.4.8 Schulausbildung (schule)

Variablenlabel	Schulausbildung
Variablenname	schule
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal enthält den Schulabschluss. Je nach Quelle sind unterschiedliche Ausprägungen möglich.</p> <p>1) BeH Mit der Umstellung des Tätigkeitsschlüssels ändern sich die möglichen Ausprägungen (siehe Bertat et al., 2013). Davor gelten:</p> <ul style="list-style-type: none"> 5 Volks-/Hauptschule o./m. Abschluss, Mittlere Reife oder gleichwertige Schulbildung 8 Fachhochschulreife oder Abitur/Hochschulreife 9 Abitur/Hochschulreife <p>Danach gelten:</p> <ul style="list-style-type: none"> 1 Kein Hauptschulabschluss 4 Hauptschulabschluss/Volksschulabschluss 6 Mittlere Reife 8 Fachhochschulreife oder Abitur/Hochschulreife <p>2) ASU, XASU, MTH Für diese Quellen sind die folgenden Ausprägungen möglich:</p> <ul style="list-style-type: none"> 1 Kein Hauptschulabschluss 4 Hauptschulabschluss/Volksschulabschluss 6 Mittlere Reife 7 Fachhochschulreife 9 Abitur/Hochschulreife

	Sie sind zu Beginn der Arbeitssuche/Maßnahmeteilnahme gültig. In der Quelle XASU kann das Merkmal bei Ausbildungsplatzsuchenden auch den angestrebten Abschluss enthalten.
Qualitätshinweis	Der Füllgrad nimmt in der BeH über die Jahre kontinuierlich ab und liegt in jüngeren Jahren unter 2/3. In der XASU nimmt er kontinuierlich zu und liegt seit 2012 über 2/3. In ASU und MTH ist der Füllgrad insgesamt gut.

6.5 Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

6.5.1 Tagesentgelt/ täglicher Leistungssatz (tentgelt)

Variablenlabel	Tagesentgelt, täglicher Leistungssatz
Variablenname	tentgelt
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>1) BeH</p> <p>In BeH-Sätzen weist das Merkmal jeweils das Bruttotagesentgelt eines Beschäftigten aus. Es wird berechnet aus dem vom Arbeitgeber gemeldeten Zeitraumentgelt sowie der Dauer des Zeitraums der (ungesplitteten) Originalmeldung in Kalendertagen. Das Tagesentgelt ist in Euro angegeben.</p> <p>Bis 1998 meldeten Arbeitgeber grundsätzlich nur das sozialversicherungspflichtige Arbeitsentgelt. Entgelte, die unterhalb der Geringfügigkeitsgrenze lagen, wurden nicht gemeldet. Entgelte, welche die Beitragsbemessungsgrenze in der gesetzlichen Rentenversicherung überschritten, werden nur bis zu dieser Grenze gemeldet. In der gesetzlichen Rentenversicherung gibt es zwei Beitragsbemessungsgrenzen. Die Grenze der Knappschaftlichen Rentenversicherung liegt meist höher als die Grenze der Arbeiter- und Angestelltenrentenversicherung. Jedoch ist eine Unterscheidung der beiden Versicherungsträger in den Daten nicht möglich.</p> <p>Seit der Einbeziehung der geringfügig Beschäftigten in das Meldeverfahren zum 01. April 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.</p> <p>Dies dürfte in der Regel auf Jahressonderzahlungen zurückzuführen sein, die der Arbeitgeber bei den Jahres-, Unterbrechungs- oder Abmeldungen zum normalen Entgelt hinzurechnen kann. Dabei ist es dann unerheblich, ob durch diese Addition die für den Meldezeitraum maßgebliche Beitragsbemessungsgrenze in der Rentenversicherung überschritten wird. Es könnte sich allerdings auch um fehlerhafte Angaben bei der Beschäftigungszeit handeln. (Die Entgelt-Informationen hingegen können auf Grund ihrer Versicherungsrelevanz als weniger fehleranfällig gesehen werden.)</p>

	<p>Die Geringfügigkeits- und Beitragsbemessungsgrenzen unterscheiden sich von Jahr zu Jahr und in West- und Ostdeutschland (maßgeblich ist jeweils der Standort des Betriebs). Eine Übersicht über diese Grenzen findet sich unter http://fdz.iab.de.</p> <p>Eine Angabe von 0 Euro ist auf "Unterbrechungsmeldungen" zurückzuführen. In diesen Zeiten besteht das Beschäftigungsverhältnis zwar rechtlich fort, es wird aber kein Entgelt gezahlt. Dies ist der Fall bei Krankheit nach Ende der Lohnfortzahlung, Mutterschaftszeiten und Sabbaticals.</p> <p>Das Tagesentgelt wird mit zwei Nachkommastellen ausgewiesen. Alle Werte größer als 0 und kleiner als 0,01 wurden auf 0,01 aufgerundet. Somit können die o.a. Unterbrechungsmeldungen mit der Bedingung Tagesentgelt = 0 identifiziert werden.</p> <p>2) LeH</p> <p>Bei LeH-Datensätzen enthält das Merkmal den täglichen Leistungssatz, jeweils umgerechnet in Euro. Es ist zu berücksichtigen, dass der tägliche Leistungssatz bei Datensätzen mit einem Original-Beginndatum, das vor dem 01. Januar 1998 liegt, für Werkzeuge und bei Datensätzen mit einem Original-Beginndatum ab 01. Januar 1998 für Kalendertage gilt.</p> <p>Die Angabe von 0 Euro ist seit 01. Januar 2005 auf Sperrzeiten oder Leistungsunterbrechungen zurückzuführen. Ist für einen Satz mit Leistungssatz gleich 0 ein Beendigungsgrund angegeben, so handelt es sich um eine Unterbrechungsmeldung. Bei Sätzen, die eine Sperrzeit widerspiegeln, ist der Anspruch derselbe wie vor Beginn der Sperrzeit.</p>
--	---

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

Variablenlabel	Beruf – ausgeübte/letzte Tätigkeit (KIdB 1988)
Variablenname	beruf
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>1) BeH</p> <p>Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V).</p> <p>Der Arbeitgeber verschlüsselt die Tätigkeit eines Beschäftigten hierbei gemäß der „Klassifizierung der Berufe 1988“ (Bundesanstalt für Arbeit, 1988), in der ca. 25.000 Berufsbezeichnungen zu finden sind. Die Berufsordnung besteht aus einer 3-stelligen Kennziffer und umfasst rund 330 Ausprägungen.</p> <p>Beschäftigungsmeldungen mit einem Enddatum nach dem 30. November 2011 werden nach dem neuen Tätigkeitsschlüssel 2010 (KIdB2010) gemeldet. Diese Werte werden durch einen Schwerpunktumstieg auf die KIdB1988 zurückgeschlüsselt. Dabei kommt es zu Unschärfen.</p>

	<p>2) ASU, XASU, MTH Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel s. 1).</p>
Qualitätshinweis	<p>Im Jahr 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben. In der XASU ist das Merkmal Beruf nahezu für den gesamten zur Verfügung stehenden Zeitraum nicht gefüllt.</p>

6.5.3 Berufsgruppe – ausgeübte/letzte Tätigkeit (KIdB 2010) (beruf2010_3)

Variablenlabel	Berufsgruppe – ausgeübte/letzte Tätigkeit (KIdB 2010), 3-Steller
Variablenname	beruf2010_3
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>1) BeH Die Berufsbezeichnung der im Meldezeitraum von der Beschäftigten oder dem Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für eine Beschäftigte oder einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V). Der Arbeitgeber verschlüsselt die Tätigkeit der Beschäftigten oder des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Die weniger detaillierte Berufsgruppe wird durch die ersten drei Stellen der Kennziffer erfasst. Das in der KIdB2010 an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt.</p> <p>Beschäftigungsmeldungen mit einem Enddatum vor dem 30. November 2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KIdB 1988) gemeldet. Diese Werte werden durch einen Schwerpunktumstieg auf die KIdB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten.</p> <p>2) LeH, ASU, XASU, MTH Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel s. 1).</p>
Qualitätshinweis	In 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.

6.5.4 Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KIdB 2010) (beruf2010_4)

Variablenlabel	Berufsuntergruppe – ausgeübte/letzte Tätigkeit (KIdB 2010), 4-Steller
Variablenname	beruf2010_4
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

Herkunft	BeH, LeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>1) BeH Die Berufsbezeichnung der im Meldezeitraum von der Beschäftigten oder dem Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Treffen mehrere Berufsbezeichnungen mit verschiedenen Schlüsselzahlen für einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V).</p> <p>Der Arbeitgeber verschlüsselt die Tätigkeit der oder des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Die weniger detaillierte Berufsuntergruppe wird durch die ersten vier Stellen der Kennziffer erfasst. Das in der KldB2010 an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt.</p> <p>Beschäftigungsmeldungen mit einem Enddatum vor dem 30. November 2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet. Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten.</p> <p>2) LeH, ASU, XASU, MTH Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel s. 1).</p>
Qualitätshinweis	Im Jahr 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird als Berufsangabe gemäß KldB2010 lediglich die Berufsgruppe (beruf2010_3) bereitgestellt.

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

Variablenlabel	Anforderungsniveau - ausgeübte/letzte Tätigkeit (KldB 2010)
Variablenname	niveau
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine

Detailbeschreibung	<p>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 einen Beschäftigten zu, ist der Arbeitgeber gehalten, diejenige Berufsbezeichnung zu wählen, welche die überwiegend ausgeübte Tätigkeit kennzeichnet (vgl. Bundesagentur für Arbeit, 2005, S. V).</p> <p>Der Arbeitgeber verschlüsselt die Tätigkeit der oder des Beschäftigten hierbei gemäß der „Klassifikation der Berufe 2010“ (Bundesagentur für Arbeit, 2011). Die Berufsgattung besteht aus einer 5-stelligen Kennziffer und umfasst rund 1.300 Ausprägungen. Das an der fünften Stelle erfasste Anforderungsniveau einer Tätigkeit wird gesondert in der Variable Anforderungsniveau (niveau) zur Verfügung gestellt.</p> <p>Beschäftigungsmeldungen mit einem Enddatum vor dem 30. November 2011 werden nach dem alten Tätigkeitsschlüssel 1988 (KldB 1988) gemeldet. Diese Werte werden durch einen Schwerpunktumstieg auf die KldB2010 umgeschlüsselt. Da der neue Tätigkeitsschlüssel wesentlich detaillierter als der alte ist, kommt es dabei zu erheblichen Unschärfen. Dies ist bei Analysen zu beachten.</p> <p>2) LeH, ASU, XASU, MTH Das Merkmal enthält den Beruf der letzten Tätigkeit. Zum Schlüssel s. 1).</p>
Qualitätshinweis	Im Jahr 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.

6.5.6 Teilzeit (teilzeit)

Variablenlabel	Teilzeit
Variablenname	teilzeit
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Die berufliche Stellung einer Beschäftigten oder eines Beschäftigten im Meldezeitraum wird vom Arbeitgeber im Rahmen der „Angaben zur Tätigkeit“ übermittelt.</p> <p>Das Merkmal „Berufsstellung“ unterscheidet zwischen Vollzeit- und Teilzeitbeschäftigten. Maßgeblich ist dabei das Verhältnis zwischen der arbeitsvertraglich vereinbarten und der betriebsüblichen Arbeitszeit.</p>
Qualitätshinweis	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.

6.5.7 Erwerbsstatus (erwstat)

Variablenlabel	Erwerbsstatus
Variablenname	erwstat

Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal nimmt für jede Datenquelle unterschiedliche Werte mit unterschiedlichen Bedeutungen an.</p> <p>1) BeH</p> <p>Für BeH-Sätze entspricht das Merkmal Erwerbsstatus der im neuen Meldeverfahren (DEÜV) ab dem 01. Januar 1999 erfassten Personengruppe. Sie weist beitrags- oder leistungsrechtliche Besonderheiten des Beschäftigungsverhältnisses aus.</p> <p>Treffen für eine Beschäftigungsmeldung mehrere Schlüssel zu, ist vom meldenden Arbeitgeber der kleinste anzugeben. Mehrheitlich handelt sich dabei um sozialversicherungspflichtige Beschäftigungsverhältnisse ohne Besonderheiten, die unter Schlüsselnummer 101 geführt werden. Entsprechend ist es möglich, dass diese Beschäftigungen leicht überschätzt werden.</p> <p>Im Meldeverfahren ist geregelt, dass Wechsel im Erwerbsstatus - z.B. die Übernahme eines Auszubildenden nach dem Ende der Berufsausbildung im Ausbildungsbetrieb - durch eine erneute Meldung erfolgen müssen (vgl. Deutsche BKK, 2016, S. 35).</p> <p>Die Personengruppe kann in Beschäftigungsmeldungen, die sich auf die Jahre vor 1999 beziehen, aber erst 1999 oder später eingegangen sind, enthalten sein. Für Meldungen, die vor 1999 eingegangen sind, wird versucht, anhand bestimmter Regeln mit Hilfe der Merkmale „Ausbildung“, „Stellung im Beruf und Arbeitszeit“ und „Beruf“ sowie weiterer Informationen eine Zuordnung zu den Personengruppen durchzuführen. In vielen Fällen ist das jedoch nicht eindeutig möglich.</p> <p>Seit dem 01. April 1999 werden im Meldeverfahren DEÜV auch geringfügig Beschäftigte erfasst.</p> <p>Diese Personengruppe lässt sich über die Ausprägungen 109 und 209 abgrenzen. Für geringfügig Beschäftigte konnten keine Daten vor dem Zeitpunkt der Einführung der Meldepflicht im Jahr 1999 zurückgeschrieben werden.</p> <p>2) LeH</p> <p>Für LeH-Sätze ist im Merkmal Erwerbsstatus die gruppierte Leistungsart angegeben. Es kann also unterschieden werden, ob eine Person Arbeitslosengeld, Arbeitslosenhilfe oder Unterhaltsgeld bezieht oder ob Beiträge zur privaten Pflegeversicherung von der BA übernommen werden.</p> <p>3) ASU/XASU</p> <p>Für ASU-Sätze zeigt das Merkmal Erwerbsstatus den Status der Arbeitssuche an. Als Ratsuchende werden Empfängerinnen und Empfänger von Arbeitslosengeld (bzw. Arbeitslosengeld II), die Leistungen unter den erleichterten Voraussetzungen des § 428 SGB III (bzw. des § 65 Abs. 4 SGB II) beziehen, sowie Nichtleistungsempfängerinnen und -empfänger ohne Vermittlungsbereitschaft i. S. des § 252 Abs. 8 SGB VI, wenn das 58. Lebensjahr vollendet wurde, geführt.</p>

	<p>Der Status "ohne Status" (Statistik: „nicht gesetzt“) impliziert hauptsächlich Personen, denen nach §10 SGB II eine Aktivierung bzw. Vermittlung nicht zuzumuten ist. Ebenso werden Personen mit einer Arbeitsunfähigkeit von mehr als 42 Tagen, die weiterhin ALG II beziehen, unter diesem Status im System geführt.</p> <p>In XASU-Sätzen enthält das Merkmal Erwerbsstatus bislang die Ausprägungen „nicht arbeitslos, aber arbeitsuchend“ sowie „arbeitsuchend und gleichzeitig arbeitslos“.</p> <p>Da die Meldelogik in vielen Fällen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben durch das in Kapitel 3.1.8 beschriebene Anonymisierungsverfahren verändert.</p> <p>4) LHG</p> <p>Für LHG-Sätze zeigt das Merkmal Erwerbsstatus an, ob die Person minderjährig erwerbsfähig, volljährig erwerbsfähig oder nicht erwerbsfähig ab Altersrentengrenze geführt wird.</p> <p>Da die Meldelogik in vielen Fällen eine Reidentifikation des exakten Geburtsdatums ermöglichen würde, wurden die ursprünglichen Datumsangaben durch das in Kapitel 3.1.8 beschriebene Anonymisierungsverfahren verändert.</p> <p>5) MTH</p> <p>Für MTH-Sätze zeigt das Merkmal Erwerbsstatus die Maßnahmenart-Kategorie an. Dies ist die höchste Hierarchiestufe der Maßnahmenart-Systematiken der BA.</p>
Qualitätshinweis	<p>In der LHG ist zu beobachten, dass es überdurchschnittlich viele 15-jährige und in geringerem Maß 16- und 17-jährige mit Hinweis auf fehlende Erwerbsfähigkeit gibt. 15- bzw. 16-jährige Hilfebedürftige der zugelassenen kommunalen Träger sind deshalb möglicherweise unterrepräsentiert, da „Nichterwerbsfähige“ nicht in die IEB übernommen werden.</p>

6.5.8 Gleitzone (gleitz)

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

6.5.9 Leiharbeit (leih)

Variablenlabel	Leiharbeit
Variablenname	leih
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal gibt an, ob es sich bei der Beschäftigung der Person um ein Leiharbeitsverhältnis handelt. Das Merkmal entstammt dem Tätigkeitsschlüssel 2010 und ist nur für Meldungen mit einem Enddatum nach dem 30. November 2011 gefüllt.
Qualitätshinweis	Im Jahr 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.

6.5.10 Befristung (befrist)

Variablenlabel	Befristung
Variablenname	befrist
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	Das Merkmal gibt an, ob es sich bei der Beschäftigung der Person um ein befristetes oder unbefristetes Arbeitsverhältnis handelt. Das Merkmal entstammt dem Tätigkeitsschlüssel 2010 und ist nur für Meldungen mit einem Enddatum nach dem 30. November 2011 gefüllt.
Qualitätshinweis	Im Jahr 2011 kommt es aufgrund der Meldeumstellung zu einer deutlichen Häufung der fehlenden Angaben.

6.5.11 Grund Abgabe Beschäftigungsmeldung / Grund Ende Leistungsbezug (grund)

Variablenlabel	Abmeldegrund / Abgabegrund / Beendigungsgrund
Variablenname	grund
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, LHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>1) BeH</p> <p>In BeH-Meldungen kennzeichnet der Abgabegrund den Anlass, aus dem der Arbeitgeber den Sozialversicherungsträgern die betreffende Beschäftigungsmeldung übermittelt hat. Allerdings werden in die IEB nicht alle im Rahmen des Meldeverfahrens mögliche Abgabegründe integriert. So beinhalten die IEB nur Meldungen mit Entgeltangaben (also Jahres-, Unterbrechungs- und Abmeldungen), während Anmeldungen wegen fehlender Entgeltinformationen nicht enthalten sind.</p>

	<p>Ein Informationsverlust ist damit jedoch nicht verbunden, da die Angaben aus einer Anmeldung mit der darauffolgenden Jahres-, Unterbrechungs- oder Abmeldung erneut übermittelt werden.</p> <p>Die Abgabegründe für Beschäftigungsmeldungen sind nach den Regeln des seit 01. Januar 1999 geltenden Meldeverfahrens (gemäß DEÜV) kodiert.</p> <p>2) LeH</p> <p>In LeH-Meldungen gibt das Merkmal den Grund für das Ende des Bezugs von Arbeitslosengeld, -hilfe oder Unterhaltsgeld an. Über die Gründe für den Beginn des Leistungsbezugs finden sich in der LeH keine Informationen, da die LeH mit den Meldungen der Arbeitsagenturen an die Krankenkassen über abgeschlossene Leistungsbezugsdauern gefüllt wird.</p> <p>3) LHG</p> <p>Bei LHG-Datensätzen ist der SGB-II-Einstellungsgrund enthalten und gibt den Grund für die Beendigung von laufenden Leistungen an. Der SGB-II-Beendigungsgrund bezieht sich auf die Person, nicht aber auf die Bedarfsgemeinschaft. Wird bei einem anderen als dem betrachteten Mitglied der Bedarfsgemeinschaft der SGB-II-Leistungsbezug beendet, werden neue Sätze für alle Mitglieder in einer Bedarfsgemeinschaft zu diesem Datum begonnen, der SGB-II-Beendigungs-/Einstellungsgrund ist jedoch nur für die betroffene Person gefüllt. Das Merkmal ist genau zum Ende des Originalsatzes gültig.</p> <p>4) ASU</p> <p>Bei ASU-Sätzen ist der Abmelde- bzw. Abgangsgrund enthalten. Im Fall eines Rechtskreiswechsels wird der Satz künstlich gesplittet und als Abmeldegrund "DS durch Splittung erzeugt" eingetragen. Zur korrekten Abbildung der Abgangsgründe ist außerdem das Merkmal „Status nach Arbeitssuche“ zu beachten.</p> <p>Ab dem 26. April 2003 erfolgte eine Reduktion der Merkmalsausprägungen. Für Auswertungen über längere Zeiträume können die alten Werte auf die aktuell gültigen anhand folgender Tabelle umgeschlüsselt werden:</p> <table border="1" data-bbox="496 1384 1104 1630"> <thead> <tr> <th>alt → neu</th> <th>alt → neu</th> <th>alt → neu</th> <th>alt → neu</th> </tr> </thead> <tbody> <tr><td>29 → 60</td><td>36 → 61</td><td>44 → 74</td><td>51 → 74</td></tr> <tr><td>30 → 60</td><td>37 → 66</td><td>45 → 77</td><td>52 → 76</td></tr> <tr><td>31 → 61</td><td>38 → 66</td><td>46 → 67</td><td>53 → 68</td></tr> <tr><td>32 → 60</td><td>39 → 71</td><td>47 → 67</td><td>54 → 78</td></tr> <tr><td>33 → 60</td><td>40 → 69</td><td>48 → 78</td><td></td></tr> <tr><td>34 → 60</td><td>42 → 65</td><td>49 → 69</td><td></td></tr> <tr><td>35 → 60</td><td>43 → 70</td><td>50 → 75</td><td></td></tr> </tbody> </table> <p>5) XASU</p> <p>Bei XASU-Sätzen ist der Abmelde- bzw. Abgangsgrund enthalten. Im Fall eines Rechtskreiswechsels wird der Satz künstlich gesplittet und als Abmeldegrund "DS durch Splittung erzeugt" eingetragen.</p>	alt → neu	alt → neu	alt → neu	alt → neu	29 → 60	36 → 61	44 → 74	51 → 74	30 → 60	37 → 66	45 → 77	52 → 76	31 → 61	38 → 66	46 → 67	53 → 68	32 → 60	39 → 71	47 → 67	54 → 78	33 → 60	40 → 69	48 → 78		34 → 60	42 → 65	49 → 69		35 → 60	43 → 70	50 → 75	
alt → neu	alt → neu	alt → neu	alt → neu																														
29 → 60	36 → 61	44 → 74	51 → 74																														
30 → 60	37 → 66	45 → 77	52 → 76																														
31 → 61	38 → 66	46 → 67	53 → 68																														
32 → 60	39 → 71	47 → 67	54 → 78																														
33 → 60	40 → 69	48 → 78																															
34 → 60	42 → 65	49 → 69																															
35 → 60	43 → 70	50 → 75																															
Qualitätshinweis	Der Anteil valider Werte (Füllgrad) des Abgabegrundes ist in der Quelle LHG über alle Jahre sehr niedrig (< 20%). Auch in der XASU ist der Füllgrad gering, jedoch verbessert er sich im Zeitablauf.																																

6.5.12 Erwerbsstatus vor Arbeitsuche (estatvor)

Variablenlabel	Erwerbsstatus vor Arbeitsuche
----------------	-------------------------------

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

6.5.13 Status nach Arbeitsuche (estatnach)

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

6.5.14 Profillage (profil)

Variablenlabel	Profillage
Variablenname	profil
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, MTH
Datentyp	numerisch
Hierarchie	keine

Detailbeschreibung	<p>Das Merkmal gibt die der Kundin oder dem Kunden beim Profiling zugewiesene Profillage wieder. Das Profiling dient der Erstellung eines Kundenprofils, d. h. einer Auflistung der arbeitsmarktrelevanten Fähigkeiten, Erfahrungen und Interessen, so dass die Verortung der Kundin oder des Kunden auf dem Arbeitsmarkt erleichtert wird. Gegen Ende des Profilings wird zusammenfassend eine Profillage festgestellt. Hierzu wird zunächst die allgemeine Integrationsprognose der Kundin oder des Kunden bestimmt. Zur Auswahl stehen:</p> <ul style="list-style-type: none"> • integrationsnah (Integration in den 1. Arbeitsmarkt innerhalb von 12 Monaten realistisch) • komplex (Integration in den 1. Arbeitsmarkt innerhalb von 12 Monaten nicht realistisch) • sonstiges <p>Von der Bestimmung der Integrationsprognose ist die Zuweisung der Profillage abhängig. Den Kundinnen und Kunden mit integrationsnaher Prognose können die Profillagen "Marktprofil", "Aktivierungsprofil" und "Förderprofil" zugewiesen werden, während die Kundinnen und Kunden mit komplexer Prognose die Profillagen "Entwicklungsprofil", "Stabilisierungsprofil" oder "Unterstützungsprofil" erhalten. Die Auswahl der konkreten Profillage richtet sich nach dem vom Vermittler angenommenen Handlungsbedarf. Wenn eine "sonstige" Prognose vorliegt, kann bei der Profillage "Zuordnung nicht erforderlich" oder - nur für SGB-II-Kundinnen und Kunden - "Integriert, aber hilfebedürftig" gewählt werden.</p>
Qualitätshinweis	Das Merkmal wurde 2009 eingeführt, aber bis 2006 aus anderen Merkmalen generiert. Die Qualität wird von der BA-Statistik seit 2010 als verlässlich eingestuft.

6.5.15 Art der Kündigung der letzten Tätigkeit (art_kuend)

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

6.5.16 Arbeitszeit des Stellengesuchs (arbeitszeit)

Variablenlabel	Arbeitszeit des Stellengesuchs
Variablenname	arbeitszeit
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU, MTH
Datentyp	Numerisch
Hierarchie	keine

Detailbeschreibung	Arbeitsuchende geben beim Vermittlungsvorgang an, welchem zeitlichen Umfang die gesuchte Stelle entsprechen sollte.
--------------------	---

6.5.17 Restanspruch/geplante Dauer (restanspruch)

Variablenlabel	Restanspruch/geplante Dauer
Variablenname	restanspruch
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LeH, MTH
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	<p>Je nach Datenquelle hat das Merkmal eine unterschiedliche Bedeutung.</p> <p>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. Januar 1998, wird der Restanspruch in Werktagen angegeben. Danach erfolgt die Angabe in Kalendertagen. Die Angabe bezieht sich nicht auf das Enddatum der Episode, sondern auf das Ende des Originalzeitraums.</p> <p>2) MTH Das Merkmal enthält die geplante Dauer der Maßnahme.</p>

6.5.18 Trägerart (traeger)

Variablenlabel	Trägerart
Variablenname	traeger
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal enthält die Trägerschaftsart des Trägers der Grundsicherung (LHG), des Trägers der Maßnahme (MTH) oder des Trägers der das Bewerberprofil verwaltet (ASU/XASU). Die Dimension enthält neben der SGB-III-Trägerschaftsart "Bundesagentur für Arbeit (BA)" drei Trägerschaftsarten des SGB II. Für weitere Informationen siehe Kapitel 2.5. Die Trägerschaftsart von Grundsicherungsträgern kann im Zeitverlauf wechseln (z.B. von ARGE zu gAw oder gE zu zkT).</p>

	<p>Zu beachten ist außerdem, dass die Grenzen der Träger in Bezug auf Kreise nicht immer trennscharf sind und somit nicht mit den Grenzen der Kreise übereinstimmen müssen. Gleiches gilt für die Arbeitsagenturen.</p> <p>Bei ASU- und XASU-Sätzen ist die Art des Trägers enthalten, der das Bewerberangebot führt.</p>
--	---

6.5.19 Beginndatum der Arbeitslosigkeit (alo_beg)

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

6.5.20 Dauer der Arbeitslosigkeit (alo_dau)

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

6.6 Ortsangaben

6.6.1 Wohnort - Kreis (wo_kreis)

Variablenlabel	Wohnort - Kreis
----------------	-----------------

Variablenname	wo_kreis
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Bundesland Kreis
Detailbeschreibung	<p>In BeH- und LeH-Sätzen liegt der Kreis des Wohnortes erst für die Jahre ab 1999 vor. Darin ist jeweils der Kreis (kreisfreie Stadt oder Landkreis) angegeben, in dem die oder der Versicherte wohnt. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des Bundeslandes, in der ersten bis dritten Stelle den Regierungsbezirk und in der ersten bis fünften Stelle den Kreis. Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null.</p> <p>In der BeH wird der Wohnort jeweils zum Jahresende ermittelt und dann einheitlich für alle Sätze eines Jahres zugespielt. Für die Quellen LHG und XASU gilt der Wohnort für den Zeitraum der Originalbeobachtung. Bei ASU, LeH und MTH enthält das Merkmal den Wohnort zum Beginn des Originalzeitraums.</p> <p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben für alle Quellen auf den Gebietsstand vom 31. Dezember 2014 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31. Dezember 2014 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 betreffende Versicherte umgezogen ist.</p>
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird als Regionalangabe lediglich das Bundesland (wo_bula) bereitgestellt.

6.6.2 Wohnort - Bundesland (wo_bula)

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

6.6.3 Wohnort - Arbeitsagentur (wo_aa)

Variablenlabel	Wohnort - Arbeitsagentur
Variablenname	wo_aa

Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, ASU, XASU, MTH
Datentyp	Numerisch
Hierarchie	Regionaldirektion Arbeitsagentur
Detailbeschreibung	<p>Das Merkmal enthält ab 1999 die Arbeitsagentur, die am Wohnort der oder des Beschäftigten/der Kundinnen und Kunden zuständig ist. Sie wird aus der Wohnadresse ermittelt. Für die Quellen LHG und XASU gilt der Wohnort für den Zeitraum der Originalbeobachtung. Bei LeH, ASU und MTH enthält das Merkmal den Wohnort zum Beginn der Meldung. Es ist für die BeH und LeH erst ab 1999 vorhanden.</p> <p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Arbeitsagenturangaben für alle Quellen auf den Gebietsstand vom 31. Dezember 2014 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Arbeitsamtsbezirk nach den Grenzen, welche die Arbeitsagenturbezirke zum 31. Dezember 2014 hatten.</p> <p>Da sich die Grenzen der Arbeitsagenturbezirke im Zeitverlauf geändert haben, würden ohne diese Gebietsstandsaktualisierung Fälle auftreten, bei denen der Arbeitsagenturbezirk einer Person wechselt, ohne dass diese Person umgezogen ist. Einen Problemfall stellt hinsichtlich der Gebietsstandsaktualisierung allerdings Berlin dar: Die Berliner Arbeitsagenturbezirke wurden im Laufe der Jahre immer wieder neu zugeschnitten, was auch durch die Gebietsstandsumschlüsselung nicht vollständig bereinigt werden konnte.</p> <p>Vor 1999 enthält das Merkmal die Arbeitsagentur, die den Leistungsfall zuletzt bearbeitet hat. Sie wird nicht aus der Wohnortadresse ermittelt und ist nur für die Quelle LeH gefüllt. Außerdem ist eine Gebietsstandsaktualisierung nicht möglich. Dies ist bei Berechnungen im Zeitverlauf (z. B. regionale Arbeitslosenzahlen) zu beachten.</p>
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird lediglich das Gebiet der Regionaldirektion, in dem der Wohnort des Versicherten liegt, ausgewiesen.

6.6.4 Wohnort - Regionaldirektion (wo_rd)

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

6.7 Betriebsmerkmale

6.7.1 WZ73 3-Steller (w73_3)

Variablenlabel	WZ73 3-Steller
Variablenname	w73_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabteilung (1-Steller) Wirtschaftsgruppe (2-Steller) Wirtschaftsklasse (3-Steller)
Detailbeschreibung	<p>Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WS73 als 3-Steller aus und ist für die Jahre 1975 bis einschließlich 2002 gefüllt.</p> <p>WS73 steht für die "Klassifikation der Wirtschaftszweige 1973" (Bundesagentur für Arbeit, 2009). Mittels eines dreistelligen Zahlencodes werden 269 Wirtschaftsklassen unterschieden, wobei die erste Stelle dieses Codes die Wirtschaftsabteilung, insgesamt 10, definiert und die beiden ersten Stellen zusammen die jeweilige Gruppe, insgesamt 95, festlegen.</p> <p>Jedem Betrieb ist nur eine Kennziffer zugeordnet. Die Zuordnung zu der entsprechenden Wirtschaftsklasse erfolgt unter Beachtung der institutionellen Ausrichtung des Betriebes.</p>

6.7.2 WZ93 5-Steller (w93_5)

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

Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w93_3) bereitgestellt.
----------------	---

6.7.3 WZ93 3-Steller (w93_3)

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

6.7.4 WZ03 5-Steller (w03_5)

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

Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller (w03_3) bereitgestellt.
----------------	---

6.7.5 WZ03 3-Steller (w03_3)

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

6.7.6 WZ08 5-Steller (w08_5)

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

6.7.7 WZ08 3-Steller (w08_3)

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

6.7.8 W73_3 vervollständigt durch Extrapolation/Imputation (w73_3_gen)

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

6.7.9 Art der Vervollständigung w73_3 (group_w73_3)

Variablenlabel	Art der Vervollständigung w73_3
Variablenname	group_w73_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine

Detailbeschreibung	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).
--------------------	---

6.7.10 W93_3 vervollständigt durch Extrapolation/Imputation (w93_3_gen)

Variablenlabel	W93_3 vervollständigt durch Extrapolation/Imputation
Variablenname	w93_3_gen
Kategorie	Betriebsmerkmale
Herkunft	BeH
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- bzw. fortgeschrieben oder mit Hilfe von Umschlüsselungstabellen ersetzt, sodass das Merkmal eine zeitkonsistente Wirtschaftszweigangabe auf Basis der Wirtschaftszweigklassifikation WZ93 enthält. Eine ausführliche Beschreibung dazu findet sich bei Eberle et al. (2011). Weitere Informationen zur Klassifikation WZ93 finden sich bei der Variablenbeschreibung w93_3.

6.7.11 Art der Vervollständigung w93_3 (group_w93_3)

Variablenlabel	Art der Vervollständigung w93_3
Variablenname	group_w93_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	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).

6.7.12 Jahr des 1. Auftretens der Betriebsnummer (grd_jahr)

Variablenlabel	Jahr des 1. Auftretens der Betriebsnummer
Variablenname	grd_jahr
Kategorie	Betriebsmerkmale
Herkunft	BeH

Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Bei dieser Variable handelt es sich um das erste Auftreten der Nummer des Betriebes im Datensatz. Wird eine Betriebsnummer in Westdeutschland erst nach 1975 bzw. in Ostdeutschland nach 1992 erstmalig nachgewiesen, könnte es sich um das Gründungsdatum des betreffenden Betriebes handeln. Allerdings könnte es sich auch um einen bereits länger existierenden Betrieb handeln, dem nach einem Wechsel der Inhaberin oder 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.

6.7.13 Erstes Auftreten Betriebsnummer (grd_dat)

Variablenlabel	Erstes Auftreten Betriebsnummer
Variablenname	grd_dat
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Bei dieser Variable 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 der Inhaberin oder des Inhabers oder einer Änderung der Rechtsform eine neue Betriebsnummer zugeordnet wurde (zur Betriebsnummernvergabe siehe Bundesagentur für Arbeit, 2007, S. 9-11). Ebenfalls könnte es sein, dass der Betrieb bereits vorher existierte, aber keine sozialversicherungspflichtig bzw. ab 1999 auch keine geringfügig Beschäftigten hatte.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen tagesgenau bereitgestellt. Ansonsten wird lediglich das Jahr des ersten Auftretens der Betriebsnummer (grd_jahr) bereitgestellt.

6.7.14 Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr)

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

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

6.7.15 Letztes Auftreten Betriebsnummer (lzt_dat)

Variablenlabel	Letztes Auftreten Betriebsnummer
Variablenname	lzt_dat
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Bei dieser Variable 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).
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen tagesgenau bereitgestellt. Ansonsten wird lediglich das Jahr des letzten Auftretens der Betriebsnummer (lzt_jahr) bereitgestellt.

6.7.16 Anzahl Beschäftigte gesamt (az_ges)

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

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

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

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

6.7.18 Anzahl geringfügig Beschäftigte (az_gf)

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

6.7.19 Mittelwert imputiertes Bruttotagesentgelt Vollzeitbeschäftigte (te_imp_mw)

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

6.7.20 Arbeitsort Kreis (ao_kreis)

Variablenlabel	Arbeitsort Kreis
Variablenname	ao_kreis
Kategorie	Ortsangaben
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Bundesland Kreis
Detailbeschreibung	<p>Angegeben ist jeweils der Kreis (kreisfreie Stadt oder Landkreis), in dem der Betrieb des Beschäftigten angesiedelt ist. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des Bundeslands, in der ersten bis dritten Stelle den Regierungsbezirk und in der ersten bis fünften Stelle den Kreis. Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null.</p> <p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben auf den Gebietsstand vom 31. Dezember 2014 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Betriebsortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31. Dezember 2014 hatten. Da sich die Grenzen der Kreise im Zeitverlauf geändert haben, würden ohne Gebietsstandsaktualisierung Fälle auftreten, bei denen die Kreiskennziffer des Betriebsortes wechselt, ohne dass dieser Betrieb seinen Standort verlagert hat.</p>
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergrößert bereitgestellt. Ansonsten wird als Regionalangabe lediglich das Bundesland (ao_bula) bereitgestellt.

6.7.21 Arbeitsort Bundesland (ao_bula)

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

6.8 Verknüpfungsmerkmale

6.8.1 Vorliegen administrativer Personendaten (match_admin)

Variablenlabel	Vorliegen administrativer Personendaten
Variablenname	match_admin
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	Generiert

Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Dieses Merkmal ist nur in der Verknüpfungsdatei enthalten. Es zeigt für jede befragte Person in den NEPS-SC6-Erhebungsdaten an, ob zu dieser Person auch Beobachtungen aus den administrativen Personendaten vorliegen. Diese Variable nimmt in wenigen Fällen den Wert 0 an, obwohl die Variable match_typ einen erfolgreichen Match anzeigt. In diesen Fällen konnte zwar ein Match durch Abgleich der Identifikatoren beider Datenquellen erreicht werden, die IEB enthält jedoch zu dieser Person keine Episoden.

6.8.2 Verknüpfungsmethode (match_typ)

Variablenlabel	Verknüpfungsart
Variablenname	match_typ
Kategorie	Verknüpfungsbezogenes Merkmal
Herkunft	Record Linkage
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Dieses Merkmal ist nur in der Verknüpfungsdatei enthalten. Es beschreibt, auf welcher Stufe des Verknüpfungsprozesses es gelungen ist, eine Person in den administrativen Daten des IAB zu identifizieren.

6.8.3 Qualität der Verknüpfung (quality)

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

7 Literatur

- Antoni, M., Ganzer, A., Berge Philipp v.** (2016). Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975-2014." FDZ-Datenreport 04/2016 (de), Nürnberg.
- Antoni, M., Schnell, R.** (2017). The past, present and future of the German Record Linkage Center (GRLC). In: Jahrbücher für Nationalökonomie und Statistik, online first.
- Antoni, M., & Seth, S.** (2012). ALWA-ADIAB–Linked individual survey and administrative data for substantive and methodological research. Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 132(1), 141-146.
- Bender, S., Hilzendegen, J., Rohwer, G., & Rudolph, H.** (1996). Die IAB-Beschäftigtenstichprobe 1975–1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197, Nürnberg.
- Bertat, T., Dundler, A., Grimm, C., Kiewitt, J., Schomaker, C., Schridde, H., & Zemann, C.** (2013). Neue Erhebungsinhalte „Arbeitszeit“, „ausgeübte Tätigkeit“ sowie „Schul- und Berufsabschluss“ in der Beschäftigungsstatistik. Methodenbericht der Statistik der Bundesagentur für Arbeit, URL: <http://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Methodenberichte/Beschaeftigungsstatistik/Generische-Publikationen/Methodenbericht-Neue-Erhebungsinhalte-Arbeitszeit-ausgeuebte-Taetigkeit-sowie-Schul-und-Berufsabschluss-in-der-Beschaeftigungsstatistik.pdf> (Zugriff: 05.04.2018).
- Bundesagentur für Arbeit.** (2005). Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung (Vol. Ausgabe Januar 2005), Nürnberg.
- Bundesagentur für Arbeit.** (2007). Handbuch für die Betriebsnummernvergabe und –pflege im Rahmen des Meldeverfahren zur Sozialversicherung (Ausgabe Dezember 2007), Nürnberg.
- Bundesagentur für Arbeit.** (2009). Klassifikation der Wirtschaftszweige 1973. Methodenbericht der Statistik der Bundesagentur für Arbeit. Nürnberg. URL: <http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Wirtschaftszweige/Klassifikation-der-Wirtschaftszweige-1973-2003/Klassifikationen-der-Wirtschaftszweige-1973-2003-Nav.html> (Zugriff: 05.04.2018)
- Bundesagentur für Arbeit.** (2011). Klassifikation der Berufe 2010–Band 1: systematischer und alphabetischer Teil mit Erläuterungen. Klassifikation der Berufe 2010, Nürnberg. URL: <http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB2010/Printausgabe-KldB-2010-Nav.html> (Zugriff: 05.04.2018)
- Bundesanstalt für Arbeit.** (1988). Klassifikation der Berufe–Systematisches und Alphabetisches Verzeichnis der Berufsbenennung, Nürnberg. URL: <http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB1975-1992/KldB1975-1992-Nav.html> (Zugriff: 05.04.2018)
- Christen, P.** (2012). Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Berlin: Springer.
- Cramer, U.** (1985). Probleme der Genauigkeit der Beschäftigtenstatistik. Allgemeines Statistisches Archiv 69: 56-68.
- Deutsche BKK.** (2016). Ratgeber Sozialversicherung 2016, Wolfsburg. URL: <http://docplayer.org/14261631-Ratgeber-sozialversicherung-2016.html> (Zugriff: 05.04.2018)

- Drews, N.** (2006). Qualitätsverbesserung der Bildungsvariable in der IAB-Beschäftigtenstichprobe 1975-2001. FDZ Methodenreport 05/2006 (de), Nürnberg.
- Eberle, J., Jacobebbinghaus, P., Ludsteck, J., & Witter, J.** (2011). Generation of time-consistent industry codes in the face of classification changes - Simple heuristic based on the Establishment History Panel (BHP). FDZ Methodenreport 05/2011 (en), Nürnberg.
- Fitzenberger, B., Osikominu, A., & Völter, R.** (2005). Imputation rules to improve the education variable in the IAB employment subsample. Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 126(3), 405-436.
- Hammon, A., Zinn, S., Aßmann, C., & Würbach, A.** (2016). Samples, Weights, and Nonresponse: The Adult Cohort of the National Educational Panel Study (Wave 2 to 6). NEPS Survey Papers.
- Kleinert, C., Matthes, B., Antoni, M., Drasch, K., Ruland, M., & Trahms, A.** (2011). ALWA—new life course data for Germany. Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 131(4), 625-634.
- Ludsteck, J., & Thomsen, U.** (2016). Imputation of the working time information for the employment register data. FDZ-Methodenreport 01/2016 (en), Nürnberg.
- Meinken, H., & Koch, I.** (2004). BA-Beschäftigtenpanel 1998-2002. Codebuch, Nürnberg.
- Paulus, W., & Matthes, B.** (2013). Klassifikation der Berufe: Struktur, Codierung und Umsteigeschlüssel. FDZ-Methodenreport 08/2013 (de), Nürnberg.
- Schild, C.-J., & Antoni, M.** (2014). Linking survey data with administrative social security data—The project „Interactions Between Capabilities in Work and Private Life “.German RLC Working Paper No. wp-grlc-2014-02.
- Schmucker, A., Seth, S., Ludsteck, J., Eberle, J., & Ganzer, A.** (2016). Betriebs-Historik-Panel 1975-2014. FDZ-Datenreport 03/2016 (de), Nürnberg.
- Schnell, R., Bachteler, T., & Bender, S.** (2004). A toolbox for record linkage. Austrian Journal of Statistics, 33(1-2), 125-133.
- Statistik der Bundesagentur für Arbeit.** (2013). Neue Erhebungsinhalte „Arbeitszeit“, „ausgeübte Tätigkeit“ sowie „Schul- und Berufsabschluss“ in der Beschäftigungsstatistik. Methodenbericht.
- Statistisches Bundesamt.** (2002). Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ93).
- Statistisches Bundesamt.** (2003). Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003).
- Statistisches Bundesamt.** (2008). Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).
- Wermter, W., & Cramer, U.** (1988). Wie hoch war der Beschäftigtenanstieg seit 1983? – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 21(4), 468-482.

8 Anhang

8.1 Auszählungen

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

9 Abkürzungsverzeichnis

ALG II	Arbeitslosengeld II
ARGE	Arbeitsgemeinschaft
ASU	Arbeitsuchendenhistorik
A2LL	Arbeitslosengeld II – Leistungen zum Lebensunterhalt
BA	Bundesagentur für Arbeit
BeH	Beschäftigtenhistorik
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 –
DÜVO	Zweite VO über die Datenübermittlung auf maschinell verwertbaren Datenträgern im Bereich der Sozialversicherung und der BA – Datenübermittlungs-Verordnung –
EDV	Elektronische Datenverarbeitung
FDZ	Forschungsdatenzentrum
FELEG	Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit
gAw	Träger mit getrennter Aufgabenwahrnehmung
gE	Gemeinsame Einrichtungen
IAB	Institut für Arbeitsmarkt- und Berufsforschung
IAB-FDZ	Forschungsdatenzentrum der Bundesagentur für Arbeit am IAB
IABS	IAB-Beschäftigtenstichprobe
IEB	Integrierte Erwerbsbiografien
ISIC	International Standard Industrial Classification of All Economic Activities
LeH	Leistungsempfängerhistorik
LHG	Leistungshistorik Grundsicherung
LifBi	Leibniz Institut für Bildungsverläufe
LifBi-FDZ	Forschungsdatenzentrum des Leibniz Institut für Bildungsverläufe
MTH	Maßnahmeteilnehmershistorik
NACE	Nomenclature générale des activités économiques dans les communautés européennes

NEPS	National Educational Panel Study
NEPS-SC6	NEPS Startkohorte 6
Schwbg	Gesetz zur Sicherung der Eingliederung Schwerbehinderter in Arbeit, Beruf und Gesellschaft – Schwerbehindertengesetz
SGB	Sozialgesetzbuch
SIAB	Stichprobe der Integrierten Arbeitsmarktbiografien
VerBIS	Vermittlungs- und Beratungsinformationssystem
XASU	Arbeitsuchendenhistorik aus XSozial-BA-SGB II
zkT	Zugelassene kommunale Träger

Impressum

FDZ-Datenreport 2/2018

Herausgeber

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

Redaktion

Dana Müller, Dagmar Theune

Technische Herstellung

Dagmar Theune

Rechte

Nachdruck - auch auszugsweise - nur mit
Genehmigung des FDZ gestattet

Bezugsmöglichkeit

http://doku.iab.de/fdz/reporte/2018/DR_02-18.pdf

Internet

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

Rückfragen zum Inhalt an:

Nadine Bachbauer
Institut für Arbeitsmarkt- und
Berufsforschung (IAB)
Forschungsdatenzentrum
Regensburger Str. 104
90478 Nürnberg

Telefon: 0911 / 179- 9579

E-Mail: Nadine.Bachbauer2@iab.de