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

FDZ·Datenreport

Dokumentation zu Arbeitsmarktdaten

01/2010
DE

Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975 - 2008

Aktualisierte Version vom 25.05.2012

Matthias Dorner
Jörg Heining
Peter Jacobebbinghaus
Stefan Seth

Bundesagentur für Arbeit

Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975 - 2008

Matthias Dorner, Jörg Heining, Peter Jacobebbinghaus, Stefan Seth
(Institut für Arbeitsmarkt- und Berufsforschung)

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 8
1.1 Einleitung 8
1.2 Datennutzung 9
1.3 Veränderungen zu früheren Versionen 10
1.4 Kurzbeschreibung 11
1.5 Merkmalsliste 13
1.6 Mengengerüst 16
2 Datenquellen 17
2.1 Beschäftigten-Historik (BeH) 17
2.2 Leistungsempfänger-Historik (LeH) 19
2.3 Leistungs-Historik-Grundsicherung (LHG/XLHG) 19
2.4 Arbeitsuchenden-Historik (ASU/XASU) 21
3 Datenaufbereitung und Stichprobenziehung 21
3.1 Bereinigungen und Aufbereitungen 21
3.1.1 Beschäftigten-Historik (BeH) 21
3.1.2 Leistungsempfänger-Historik (LeH) 22
3.1.3 Leistungs-Historik-Grundsicherung (LHG/XLHG) 23
3.1.4 Arbeitsuchenden-Historik (ASU/XASU) 23
3.1.5 Bereinigungen der Gesamtdatei 24
3.2 Episodensplitting 24
3.3 Stichprobenziehung 25
3.4 Fehlende Werte 25
4 Datenqualität und Probleme 26
4.1 Beschäftigten-Historik (BeH) 26
4.2 Leistungsempfänger-Historik (LeH) 27
4.3 Leistungs-Historik-Grundsicherung (LHG/XLHG) 27
4.4 Arbeitsuchenden-Historik (ASU/XASU) 28
4.4.1 ASU 28
4.4.2 XASU 29
5 Variablenbeschreibung 30
5.1 Identifikatoren 30
5.1.1 Systemfreie Personennummer (persnr) 30
5.1.2 Systemfreie Betriebsnummer (betnr) 30
5.2 Gültigkeitszeitraum 31
5.2.1 Beginndatum Originalsatz (begorig) 31
5.2.2 Enddatum Originalsatz (endorig) 31
5.2.3 Beginndatum der Episode (begepi) 32
5.2.4 Enddatum der Episode (endepi) 32
5.3 Informationen zur Person 32
5.3.1 Geschlecht (frau) 32
5.3.2 Geburtsjahr (gebjahr) 33
5.3.3 Staatsangehörigkeit (nation) 33
5.3.4 Staatsangehörigkeit vergröbert (nation_gr) 33
5.3.5 Familienstand (famst) 33

5.3.6 Kinderzahl (kind) 34
5.3.7 Ausbildung (bild) 34
5.3.8 Schulausbildung (schbild) 36
5.3.9 Qualifikation (quali) 36
5.3.10 Schwerbehindertenstatus (schweb) 37
5.3.11 Gesundheitliche Einschränkungen (gesund_ein) 37
5.4 Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche 37
5.4.1 Abgabegrund/ Beendigungsgrund/ SGB-II-Einstellungsgrund /Abmeldegrund (grund) 37
5.4.2 Tagesentgelt, täglicher Leistungssatz (tentgelt) 39
5.4.3 Gleitzone (gleitz) 40
5.4.4 Beruf – ausgeübte Tätigkeit (beruf) 40
5.4.5 Stellung im Beruf und Arbeitszeit (stib) 41
5.4.6 Erwerbsstatus (erwstat) 42
5.4.7 Erwerbsstatus vor Arbeitsuche (estatvor) 43
5.4.8 Leistungsart (lart) 43
5.4.9 BA-Kundengruppe (kundengr) 43
5.4.10 Art der Kündigung (art_kuend) 44
5.4.11 Bereitschaft zur bundesweiten Stellensuche (mobil) 44
5.4.12 Arbeitszeit des Stellengesuchs (arbzeit) 44
5.4.13 Restanspruchsdauer Arbeitslosengeld (restanspruch) 45
5.4.14 Trägerart (traeger) 45
5.5 Betriebsmerkmale 45
5.5.1 Wirtschaftszweig 73 – Wirtschaftsgruppe 73 (w73_3) 45
5.5.2 Wirtschaftszweig 93 – Wirtschaftsunterklasse 93 (w93_5) 46
5.5.3 Wirtschaftszweig 93 – Wirtschaftsgruppe 93 (w93_3) 46
5.5.4 Wirtschaftszweig 03 – Wirtschaftsunterklasse 03 (w03_5) 47
5.5.5 Wirtschaftszweig 03 – Wirtschaftsgruppe 03 (w03_3) 47
5.5.6 Wirtschaftszweig 08 – Wirtschaftsunterklasse 08 (w08_5) 48
5.5.7 Wirtschaftszweig 08 – Wirtschaftsgruppe 08 (w08_3) 48
5.5.8 Erstes Auftreten des Betriebes (grd_jahr) 48
5.5.9 Letztes Auftreten des Betriebes (lzt_jahr) 49
5.5.10 Anzahl der Beschäftigten gesamt (az_ges) 49
5.5.11 Anzahl Vollzeitbeschäftigte (az_ges_vz) 49
5.5.12 Anzahl geringfügig Beschäftigter (az_gf) 50
5.5.13 Betriebsentgelt (te_med) 50
5.6 Ortsangaben 50
5.6.1 Wohnort Kreis (wo_kreis) 50
5.6.2 Wohnort Arbeitsagentur (wo_aa) 51
5.6.3 Wohnort Regionaldirektion (wo_rd) 52
5.6.4 Wohnort Bundesland (wo_bula) 52
5.6.5 Arbeitsort Kreis (ao_kreis) 53
5.6.6 Arbeitsort Arbeitsagentur (ao_aa) 53
5.6.7 Arbeitsort Regionaldirektion (ao_rd) 54
5.6.8 Arbeitsort Bundesland (ao_bula) 54
5.7 Generierte technische Merkmale 54
5.7.1 Quelle des Satzes (quelle) 54

5.7.2 Satzzähler Konto (spell) 54
5.7.3 Satzzähler pro Episode (level2) 55
5.7.4 Satzzähler pro Episode und Quelle (level1) 55
5.7.5 Jahr (jahr) 56
6 Literatur 57
7 Anhang 58
7.1 Bereinigungsschritte in der IABS 58
7.2 Auszählungen 60
8 Abkürzungsverzeichnis 61

Zusammenfassung

Dieser Datenreport beschreibt die Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975 – 2008.

Abstract

This datareport describes the Sample of Integrated Labour Market Biographies (SIAB) 1975-2008.

Keywords: German administrative micro data, labour market data, data manual

Für Ihre Mitarbeit und Unterstützung danken wir den Kollegen des Forschungsdatenzentrums (FDZ), die sich in der Arbeitsgruppe zur Stichprobe der integrierten Arbeitsmarktbiografien beteiligt haben, sowie dem Servicebereich IT- und Informationsmanagement (ITM) des Instituts für Arbeitsmarkt- und Berufsforschung.

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

Die Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) ist eine 2%-Stichprobe aus der Grundgesamtheit der Integrierten Erwerbsbiografien (IEB) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB). Die IEB bestehen aus der Gesamtheit der Personen, die im Beobachtungszeitraum mindestens einmal einen der folgenden Zustände aufweisen:

- Sozialversicherungspflichtige Beschäftigung (erfasst ab 1975)
- Geringfügige Beschäftigung (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) als arbeitsuchend gemeldet (erfasst ab 2000)
- (Geplante) Teilnahme an arbeitsmarktpolitischer Maßnahme (erfasst ab 2000)

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

Während die Beschäftigten-Historik (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 Leistungsempfänger-Historik (LeH) bzw. in der Leistungs-Historik-Grundsicherung (LHG und XLHG) erfasst. Die Arbeitsuchenden-Historiken (ASU und X-ASU) stellen die Datenquelle für die durch die BA erfassten Zeiten der Arbeitsuche dar, während die (geplante) Teilnahme an arbeitsmarktpolitischen Maßnahmen in der Maßnahme-Teilnahme-Historik (MTH) erfasst werden.

Mit Ausnahme der (geplanten) Teilnahme an arbeitsmarktpolitischen Maßnahmen aus der MTH sind alle Datenquellen der IEB auch in der SIAB enthalten.

Die SIAB ersetzt die IAB-Beschäftigtenstichprobe (IABS) und die Stichprobe der Integrierten Erwerbsbiografien (IEBS). Durch die Zusammenführung der Datenquellen von IABS und IEBS kann nun ein Datensatz angeboten werden, der die Vorteile der IABS (langer Beobachtungszeitraum) mit den Vorteilen der IEBS (vollständige Erwerbsbiografien) verbindet. Die SIAB wird im Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit (BA) im IAB produziert.

Die SIAB umfasst die Erwerbsverläufe von 1.659.025 Individuen, deren Erwerbsbiografien in insgesamt 40.501.525 Datenzeilen dokumentiert sind.

Der vorliegende Datenreport beschreibt die Variablen der schwach anonymisierten Version der SIAB, welche weitgehend die Originaldaten, d.h. nicht anonymisierte Daten, beinhaltet. Lediglich die direkten Identifikatoren, wie die Sozialversicherungs- und die Betriebsnummer,

wurden durch systemfreie Nummern ersetzt. Zum Schutz der Anonymität der Betroffenen werden einige Merkmale als besonders sensibel eingestuft und nur auf gesonderten Antrag (siehe Kapitel 1.2) weitergegeben.

Der Datenreport zur SIAB gliedert sich wie folgt. Neben der Einleitung enthält Kapitel 1 unter anderem Informationen zum Datenzugang sowie eine Kurzbeschreibung der Daten, das Mengengerüst und eine Merkmalsliste. Eine Beschreibung der einzelnen Datenquellen findet sich in Kapitel 2. Die Datenaufbereitung sowie die Datenqualität werden in den Kapitel 3 und 4 diskutiert, während die einzelnen Merkmale in Kapitel 5 dargestellt werden.

1.2 Datennutzung

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

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

Die SIAB-Daten sind modular gegliedert und in mehreren Dateien abgelegt. Ein Modul, das im weiteren Verlauf Personendatei genannt wird, enthält Identifikatoren (systemfreie Personen- und Betriebsnummer), die persönlichen Merkmale, die Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche, die Merkmale zum Wohnort sowie technische Merkmale. Ein zweites Modul, die Betriebsdatei, enthält die Betriebsnummer, das Jahr und Merkmale zu Arbeitsort und Wirtschaftszweigen in vergröberter Form sowie weitere nicht-sensible Betriebsinformationen zum Stichtag 30. Juni.

Somit sind Betriebs- und Personenmerkmale organisatorisch voneinander getrennt, was die Struktur der Daten deutlich macht und Speicherplatz in der Datenhaltung spart. Die Verknüpfung von Betriebs- und Personendatei erfolgt über die Betriebsnummer und das Jahr des Datensatzes. Die in der Merkmalsliste (vgl. S. 13 ff.) mit einem „*“ gekennzeichneten Variablen, sind in der Betriebsdatei enthalten. Zu beachten ist, dass das für die Verknüpfung beider Module notwendige Merkmal „Betriebsnummer“ sowohl in der Personen- als auch in der Betriebsdatei vorhanden ist.

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

Personendatei:

- Staatsangehörigkeit (nation)
- Gesundheitliche Einschränkungen (gesund_ein)
- Schwerbehindertenstatus (schweb)
- Wohnort Arbeitsagentur (wo_aa)
- Wohnort Kreis (wo_kreis)

Betriebsdatei:

- Arbeitsort Arbeitsagentur (ao_aa)
- Arbeitsort Kreis (ao_kreis)
- Wirtschaftszweig - Wirtschaftsunterklasse (w93_5)
- Wirtschaftszweig - Wirtschaftsunterklasse (w03_5)
- Wirtschaftszweig - Wirtschaftsunterklasse (w08_5)

Diese Merkmale werden dem Nutzer in den betreffenden Dateien bereitgestellt. Auf begründeten Antrag hin können aus dem Betriebs-Historik-Panel (BHP) weitere Betriebsmerkmale bereitgestellt werden. Nähere Informationen zum BHP finden sich unter http://fdz.iab.de/.

1.3 Veränderungen zu früheren Versionen

Die SIAB ersetzt die IABS und die IEBS. Daher werden im Folgenden die wichtigsten Unterschiede zu diesen Vorgängerprodukten beschrieben.

	SIAB 1975 - 2008	**IABS 2004**	**IEBS 2008**
Erhebungsdesign	2% der Personen in den IEB	2% der **Beschäftigten** der BLH	Geburtstagsstichprobe aus IEB, 8/365 ≈ 2,2%
Quellen	Beschäftigung, Leistungsempfang, Arbeitsuche	Beschäftigung, Leistungsempfang	Beschäftigung, Leistungsempfang, Arbeitsuche, Maßnahmen
Datenbasis	Datenbasis IEB	Datenbasis BLH	Datenbasis IEB
Altersfilter	Ja. Datensätze, bei denen das Alter kleiner 13 und größer als 75 ist, sind nicht enthalten.	Kein Altersfilter	Ja. Datensätze, bei denen das Alter kleiner 13 und größer als 75 ist, sind nicht enthalten.
Jahresmeldungen mit Personengruppe 101 und Entgelt == 0	Nicht Enthalten	Enthalten	Nicht Enthalten

Tabelle 1: Unterschiede zur IABS und zur IEBS

Während die Daten in der SIAB und der IEBS weitestgehend den zugrundeliegenden Originaldaten entsprechen, wurde für die IABS eine Reihe von Bereinigungsschritten durchgeführt. Neben grundsätzlichen Bereinigungen wurden in der IABS auch Korrekturen bei mehrfachem Leistungsbezug, bei zeitlichen Überschneidungen (1 Tag) von Sätzen, bei Überschneidungen zwischen BeH- und LeH-Sätzen sowie eine Streikbereinigung durchgeführt. Ferner kam für ausgewählte Datensätze und Lücken ein Ergänzungsverfahren zum Einsatz. Der Anhang des Datenreports enthält eine detaillierte Beschreibung der genannten Korrekturen und Bereinigungen.

1.4 Kurzbeschreibung

Aktuelle Datenversion	Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975 - 2008
Art und Umfang der Daten	Schwach anonymisierte, 2%-Zufallsauswahl von Personenkonten aus den Integrierten Erwerbsbiografien (IEB) des IAB.

Kurzbeschreibung des Inhalts	**Beschäftigten-Historik (BeH):** Jahres- und Abmeldungen zur Sozialversicherung für sozialversicherungspflichtig und geringfügig Beschäftigte **Leistungsempfänger-Historik (LeH):** Informationen zu Leistungsbezug nach SGB III für Bezieher von Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld **Leistungs-Historik-Grundsicherung aus A2LL (LHG):** Daten zu Personen, die Leistungen zur Sicherung des Lebensunterhalts nach SGB II beziehen (ARGE und getrennte Trägerschaft) **Leistungs-Historik-Grundsicherung aus XSozial-BA-SGB II (XLHG):** Daten zu Personen, die Leistungen zur Sicherung des Lebensunterhalts nach SGB II beziehen (zugelassene kommunale Träger und getrennte Trägerschaft) **Arbeitsuchenden-Historik (ASU):** Informationen zur Arbeitsuche (Bewerberangebote) **Arbeitsuchenden-Historik aus XSozial-BA-SGB II (XASU):** Informationen zur Arbeitsuche, die über den Übermittlungsstandard XSozial-BA-SGBII von zugelassenen kommunalen Trägern oder getrennten Trägern an die BA gemeldet werden. **Sensible Merkmale:** Staatsangehörigkeit (nation) Gesundheitliche Einschränkungen (gesund_ein) Schwerbehindertenstatus (schweb) Wohnort Arbeitsagentur (wo_aa) Wohnort Kreis (wo_kreis) Arbeitsort Arbeitsagentur (wo_aa) Arbeitsort Kreis (ao_kreis) Wirtschaftsunterklasse (w93_5) Wirtschaftsunterklasse (w03_5) Wirtschaftsunterklasse (w08_5)
Merkmale mit regionaler Relevanz	Arbeitsort und Wohnort, Bundesländer, Kreise, Regionaldirektionen, Arbeitsagenturbezirke. (Einzelne dieser Merkmale sind nicht für alle Datenquellen und Zeiträume gefüllt.)
Gebietsstand Datum	BeH: 31.12.2008 LeH: 31.12.2008
Gebietsstand Typ	BeH, LeH: gebietsstandsbereinigt ASU, XASU, LHG, XLHG: Originalgebietsstand; keine Bereinigung
Zeitraum	Der Zeitraum hängt von der Datenquelle ab. BeH 01.01.1975 - 31.12.2008 (2006, 2007: 18-Monatsdatei; 2008: 12-Monatsdatei) LeH 01.01.1975 - 14.09.2009 ASU 01.01.2000 - 14.09.2009 LHG 01.01.2005 - 14.10.2009 XASU 01.01.2005 - 14.10.2009 XLHG 01.11.2005 - 14.10.2009

Frequenz der Aktualisierung	Laufend
Untersuchungseinheit	Sozialversicherungspflichtig Beschäftigte (ab 1999 auch geringfügig Beschäftigte), Leistungsempfänger, Arbeitsuchende
Fallzahlen	1.659.024 Personen, 40.501.525 Originalsätze, 32.419.594 überschneidungsfreie Sätze
Bereinigungen - Kurzbeschreibung	Die SIAB-Daten werden weitgehend unbereinigt aus den IEB übernommen. Es finden lediglich Aufbereitungen wie z.B. Umkodierungen von einzelnen Variablen statt.
Besonderheiten - Kurzbeschreibung	Wie die Daten in der IEB sind auch die Daten in der SIAB episodengesplittet. Dabei werden sich teilweise überschneidende Datensätze innerhalb eines Personenkontos gesplittet, sodass nur überschneidungsfreie oder vollständig parallele Zeiträume vorkommen.
Dateiorganisation	Die SIAB-Daten werden in zwei Dateien abgelegt. Die eine enthält personenbezogene und die andere betriebsbezogene Informationen.
Dateiformat, Dateigröße	STATA (Personen-Datei circa 3,5 GB), SPSS und SAS auf Anfrage
Datenzugang	Gastaufenthalt im FDZ der BA im IAB und anschließende Datenfernverarbeitung
Anonymisierungsgrad	Schwach anonymisiert
Zitierweise	Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975 - 2008, Nürnberg 2010

1.5 Merkmalsliste

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

	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: Der Schulabschluss ist lediglich für ASU- und XASU-Sätze gefüllt, die Sätze der anderen Datenquellen enthalten in dieser Variablen „.n“. Eine weitere Besonderheit der SIAB 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- bzw. XLHG-Sätze den Status SGB-II und für ASU- bzw. XASU-Sätze den Arbeitsuche-Status. Nicht bei jedem Merkmal sind diese Unterschiede direkt aus dem Variablennamen ersichtlich.

Merkmalsliste	Modul	Seite	BeH	LeH	LHG	XLHG	ASU	XASU
Identifikatoren		**30**						
Systemfreie Personennummer (persnr)		30						
Systemfreie Betriebsnummer (betnr)	*	30						
Gültigkeitszeitraum		**31**						
Beginndatum Originalsatz (begorig)		31						
Enddatum Originalsatz (endorig)		31						
Beginndatum der Episode (begepi)		32						
Enddatum der Episode (endepi)		32						
Informationen zur Person		32						
Geschlecht (frau)		32						
Geburtsjahr (gebjahr)		33						
Staatsangehörigkeit (nation)		33						
Staatsangehörigkeit vergröbert (nation_gr)		33						
Familienstand (famst)		33						
Kinderzahl (kind)		34						
Ausbildung (bild)		34						
Schulausbildung (schbild)		36						
Qualifikation (quali)		36						
Schwerbehindertenstatus (schweb)		36						
Gesundheitliche Einschränkungen (gesund_ein)		37						
Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche		**37**						
Abgabegrund/ Beendigungsgrund/ SGB-II-Einstellungsgrund /Abmeldegrund (grund)		37						
Tagesentgelt, täglicher Leistungssatz (tentgelt)		39						
Gleitzone (gleitz)		40						
Beruf – ausgeübte Tätigkeit (beruf)		40						
Stellung im Beruf und Arbeitszeit (stib)		41						
Erwerbsstatus (erwstat)		42						
Erwerbsstatus vor Arbeitsuche (estatvor)		43						
Leistungsart (lart)		43						

Merkmalsliste	Modul	Seite	BeH	LeH	LHG	XLHG	ASU	XASU
BA-Kundengruppe (kundengr)		43						
Art der Kündigung (art_kuend)		44						
Bereitschaft zur bundesweiten Stellensuche (mobil)		44						
Arbeitszeit des Stellengesuchs (arbzeit)		44						
Restanspruchsdauer Arbeitslosengeld (restanspruch)		45						
Trägerart (traeger)		45						
Betriebsmerkmale		**45**						
Wirtschaftszweig 73 – Wirtschaftsgruppe 73 (w73_3)	*	45						
Wirtschaftszweig 93 – Wirtschaftsunterklasse 93 (w93_5)	*	46						
Wirtschaftszweig 93 – Wirtschaftsgruppe 93 (w93_3)	*	46						
Wirtschaftszweig 03 – Wirtschaftsunterklasse 03 (w03_5)	*	47						
Wirtschaftszweig 03 – Wirtschaftsgruppe 03 (w03_3)	*	47						
Wirtschaftszweig 08 – Wirtschaftsunterklasse 08 (w08_5)	*	48						
Wirtschaftszweig 08 – Wirtschaftsgruppe 08 (w08_3)	*	48						
Erstes Auftreten des Betriebes (grd_jahr)	*	48						
Letztes Auftreten des Betriebes (lzt_jahr)	*	49						
Anzahl der Beschäftigten gesamt (az_ges)	*	49						
Anzahl Vollzeitbeschäftigte (az_ges_vz)	*	49						
Anzahl geringfügig Beschäftigter (az_gf)	*	50						
Betriebsentgelt (te_med)	*	50						
Ortsangaben		**50**						
Wohnort Kreis (wo_kreis)		50						
Wohnort Arbeitsagentur (wo_aa)		50						
Wohnort Regionaldirektion (wo_rd)		52						
Wohnort Bundesland (wo_bula)		52						
Arbeitsort Kreis (ao_kreis)	*	53						
Arbeitsort Arbeitsagentur (ao_aa)	*	53						

Merkmalsliste	Mo-dul	Sei-te	BeH	LeH	LHG	XLHG	ASU	XASU
Arbeitsort Regionaldirektion (ao_rd)	*	54						
Arbeitsort Bundesland (ao_bula)	*	54						
Generierte technische Merkmale		**54**						
Quelle des Satzes (quelle)		54						
Satzzähler Konto (spell)		54						
Satzzähler pro Episode (level2)		55						
Satzzähler pro Episode und Quelle (level1)		55						
Jahr (jahr)		56						

1.6 Mengengerüst

Fallzahlen der SIAB	**vor Splitting**	**nach Splitting**
BeH	24.336.041	27.056.515
LeH	3.613.235	5.036.568
LHG	621.623	1.562.609
XLHG	124.361	235.737
ASU	3.630.967	6.388.559
XASU	93.367	221.537
Sätze insgesamt	32.419.594	40.501.525
Personen	1.659.024	

Tabelle 2: Mengengerüst

2 Datenquellen

Die SIAB wurde als 2%-Stichprobe aus den Integrierten Erwerbsbiografien des IAB gezogen. Diese vereint in sich wiederum Daten aus fünf verschiedenen Datenquellen, die jeweils Informationen aus unterschiedlichen Verfahren enthalten können.

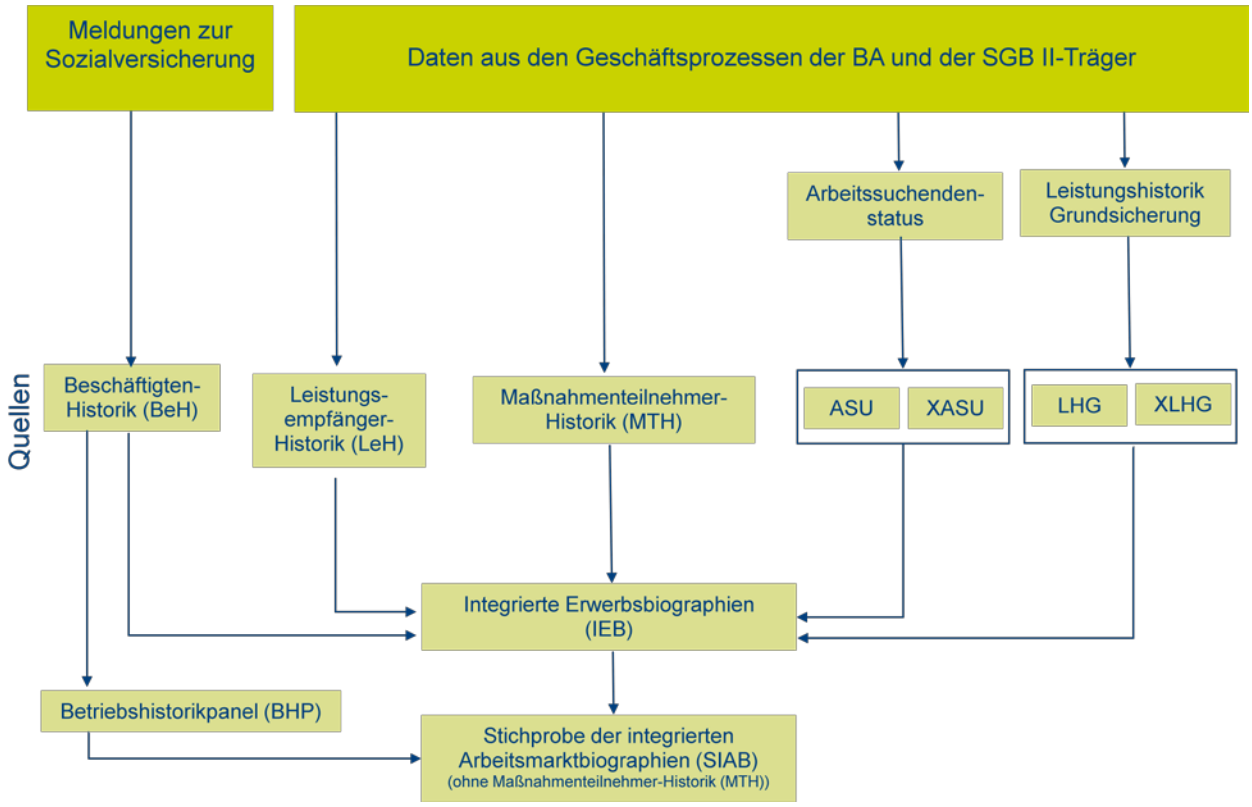

Abbildung 1: Quellen der IEB

In die SIAB werden Datensätze aus allen dargestellten Quellen mit Ausnahme der (geplanten) Teilnahmen an arbeitsmarktpolitischen Maßnahmen (MTH) aus der IEB übernommen.

2.1 Beschäftigten-Historik (BeH)

Die Quelle für Informationen über Beschäftigung ist die Beschäftigten-Historik (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. 4 ff.; Wermter/Cramer 1988). Es verlangt von den Arbeitgebern Meldungen für alle sozialversicherungspflichtig beschäftigten Arbeitnehmer an die zuständigen Sozialversicherungsträger. Die BeH erfasst alle Arbeiter und Angestellten sowie alle Auszubildenden, soweit sie nicht von der Sozialversicherungspflicht befreit sind. Dies bedeutet, dass Beamte, Selbstständige und

ordentliche Studierende (vgl. Cramer 1985) grundsätzlich nicht in der BeH erfasst werden. Seit der Änderung des Meldeverfahrens zum 1. Januar 1999 werden auch geringfügig Beschäftigte und mithelfende Familienangehörige erfasst (enthalten erst ab 1. April 1999).
Jedes Jahr, in dem ein Beschäftigungsverhältnis für eine Person existiert, ist durch mindestens eine Meldung abgebildet. Die Daten werden von den Krankenkassen aufgenommen, von der Bundesagentur für Arbeit in einer laufenden Datei gesammelt und anschließend vom IAB in seine Historik-Datei integriert. Der gegenwärtige Beobachtungszeitraum der BeH erstreckt sich vom 1. Januar 1975 bis zum 31. Dezember 2008.
Die Merkmale der SIAB Betriebs-Datei werden zum Teil dem Betriebs-Historik-Panel (BHP) entnommen, das ebenfalls auf der BeH basiert. Die SIAB Betriebs-Datei erfasst dabei folgende Merkmale:

- Arbeitsort Kreis (ao_kreis)
- Arbeitsort Bundesland (ao_bula)
- Wirtschaftszweig 73 (w73_3)
- Wirtschaftszweig 93 (w93_3 bzw. w93_5)
- Wirtschaftszweig 03 (w03_3 bzw. w03_5)
- Wirtschaftszweig 08 (w08_3 bzw. w08_5)
- Anzahl der Beschäftigten (az_ges)
- Anzahl Vollzeitbeschäftigte (az_ges_vz)
- Anzahl geringfügig Beschäftigte (az_gf)
- Betriebsentgelt (te_med)
- Gründungsjahr des Betriebes (grd_jahr)
- Letztes Auftreten des Betriebes (ltz_jahr)

Zwei Dinge müssen bei der Verknüpfung von Personen- mit Betriebsdaten berücksichtigt werden:

1. Die Merkmale in der Betriebs-Datei werden zum 30.6. eines Jahres aggregiert. Die Betriebsmerkmale sind deshalb, im Unterschied zu den Personenangaben, keine zeitraumbezogenen Daten, sondern nur am 30.6. jedes Jahres exakt gültig[1].
2. Zum anderen wird das BHP direkt aus der BeH erstellt, die SIAB Personen-Datei jedoch aus den IEB (siehe Abbildung: Quellen der IEB), in die nicht alle Datensätze der BeH übernommen werden.

Die Verknüpfung der Betriebs- an die Personen-Datei erfolgt über die programmspezifischen Kommandos der zur Datenaufbereitung und -analyse genutzten Softwarepakete. So kann

[1] Extrembeispiel: Es liegt eine Beschäftigungsmeldung vom 1.1. bis zum 30.5.2006 vor, der Betrieb geht im Juni 2006 Konkurs. Dann gibt es im BHP zu diesem Betrieb für 2006 keine Informationen.

die Verknüpfung beispielsweise in Stata unter Verwendung des „`merge`“-Befehls in Verbindung mit der entsprechenden Pfadangabe (siehe Kapitel 5.7.5) durchgeführt werden.

Nicht alle Merkmale des BHP sind standardmäßig in der SIAB enthalten. Auf begründeten Antrag hin können einzelne weitere Merkmale des BHP vom FDZ zugespielt werden. Nähere Informationen zum BHP finden sich im entsprechenden FDZ-Datenreport unter http://fdz.iab.de

2.2 Leistungsempfänger-Historik (LeH)

Die Leistungsempfängerhistorik (LeH) des IAB erfasst Zeiträume, in denen Personen Lohnersatzleistungen von der Bundesagentur für Arbeit beziehen. Die Leistungen beinhalten Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld, also keine Leistungen nach dem Rechtskreis SGB II (z.B. Arbeitslosengeld II). Der Leistungsanspruch, der in der LeH abgebildet wird, hängt von der Erfüllung gesetzlicher Voraussetzungen ab, z. B. Bedürftigkeit im Falle der Arbeitslosenhilfe oder Erfüllung der Anwartschaftszeit beim Arbeitslosengeld. Reine Arbeitslosigkeitsperioden ohne Leistungsbezug nach SGB III werden deshalb nicht in der Leistungsempfänger-Historik berichtet. Der Beobachtungszeitraum der Leistungsempfänger-Historik erstreckt sich vom 1. Januar 1975 bis zum 14. Dezember 2009.

2.3 Leistungs-Historik-Grundsicherung (LHG/XLHG)

Während die LeH die Leistungen des Rechtskreises SGB III enthält, besteht die Leistungs-Historik-Grundsicherung aus Meldungen, die Leistungsbezug nach SGB II darstellen. Leistungsbezug im Rahmen der Leistungs-Historik-Grundsicherung umfasst sowohl Leistungen zur Sicherung des Lebensunterhalts (z. B. ALG II), wie auch Zuschläge zum ALG oder auch zusätzliche Leistungen. Im Gegensatz zu den Leistungen aus dem Rechtskreis des SGB III ist die BA nicht mehr zwingend alleiniger Träger der Leistungen. Es wird demnach unterschieden nach den drei möglichen Trägerarten:

- Arbeitsgemeinschaften (ARGE) aus gemeinsamer Aufgabenwahrnehmung von BA und Kommune
- Bei getrennter Trägerschaft bleibt die Aufgabenteilung zwischen BA und Kommune bestehen[2]

[2] 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.

- Bei zugelassenen kommunalen Trägern, die auch als optierende Kommunen oder Optionskommunen bezeichnet werden, übernimmt die Kommune Aufgaben aus dem Bereich des SGB II

Die Leistungs-Historik-Grundsicherung umfasst zwei Quellen, die sich im Übermittlungsverfahren der Daten unterscheiden: Die Daten der „Leistungs-Historik-Grundsicherung aus A2LL“ (LHG) stammen aus dem BA-Verfahren A2LL, während die „Leistungs-Historik-Grundsicherung aus XLHG-BA-SGB II“ (XLHG) Daten erfasst, welche nach dem Übermittlungsstandard X-Sozial-BA-SGB II gemeldet wurden.

In der Regel findet A2LL bei ARGEn Verwendung, während der Standard X-Sozial-BA-SGB II von den optierenden Kommunen genutzt wird. Bei Kommunen mit getrennter Trägerschaft kommen beide Verfahren zum Einsatz.

Eine Besonderheit gegenüber der LHG besteht darin, dass die Ermittlung der Höhe des Leistungsbezugs nicht auf individueller Ebene erfolgt, sondern auf der Ebene von Bedarfsgemeinschaften. Eine Verknüpfung von Individuen mit SGB-II-Leistungsbezug auf Ebene der Bedarfsgemeinschaften ist in diesem Datensatz jedoch nicht möglich. In LHG und XLHG sind grundsätzlich alle gemeldeten Berechtigten nach § 7 SGB II, also erwerbsfähige Hilfebedürftige und ggf. ihre Bedarfsgemeinschaftsmitglieder enthalten. Jedoch werden nur Zeiträume übernommen, in denen eine Person erwerbsfähig oder jünger als 64 Jahre ist.

Es sind jeweils überschneidungsfreie Zeiträume der Leistungsberechtigung einer Person in einer bestimmten Bedarfsgemeinschaft abgebildet. Die Sätze werden jedoch

- an relevanten (14, 15, 18, 25 und 65) Geburtstagen der Mitglieder einer Bedarfsgemeinschaft,
- bei Veränderungen in der Zusammensetzung einer Bedarfsgemeinschaft (z.B. durch Zu-/Abgänge),
- bei Merkmalsänderungen des Bedarfsgemeinschafts-Kunden sowie
- bei Beginn und Ende eines Sanktionssachverhalts für Sätze ab dem 1. April 2006

gesplittet.

Im Hinblick auf Sanktionen ist zu beachten, dass Beginn und Ende eines Sanktionssachverhalts zwar zu einem Splitting der Sätze führen, jedoch 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.

Die Datenquelle LHG ist ab der Einführung des ALG II am 1. Januar 2005 bis zum 14. Oktober 2009 gefüllt.Dagegen sind Informationen aus der XLHG ab dem 1. November 2005 bis zum 14. Oktober 2009 enthalten.

2.4 Arbeitsuchenden-Historik (ASU/XASU)

Daten über Arbeitsuchende werden in der Arbeitsuchenden-Historik gespeichert. Die Quelle ASU enthält dabei Informationen zu Arbeitsuchenden, welche bei Arbeitsagenturen verzeichnet sind und schließt ab 2005 auch getrennte Trägerschaften mit ein. In der Quelle XASU dagegen sind ab 2005 die Daten von arbeitsuchenden Personen im ALG-II-Bezug von zugelassenen kommunalen Trägern enthalten, welche nach dem X-Sozial-Standard gemeldet werden.

3 Datenaufbereitung und Stichprobenziehung

3.1 Bereinigungen und Aufbereitungen

Bevor die Daten aus den in Kapitel 2 genannten Quellen zur IEB zusammengefügt werden, durchlaufen sie mehrere Bereinigungsschritte, welche in den folgenden Abschnitten kurz dargestellt werden.

Alle der nachfolgend genannten Bereinigungen bzw. Aufbereitungen beziehen sich auf die Rohdaten aus den einzelnen Quellen bzw. auf die Gesamtdatei der IEB, welche die Grundgesamtheit für die SIAB-Daten darstellt. Die SIAB-Daten werden nicht mehr bereinigt und entsprechen somit (nahezu) den Originaldaten der IEB. Es werden lediglich vereinzelte Umkodierungen einzelner Variablen oder aus Anonymisierungsgründen auch Vergröberungen der Daten durchgeführt.

3.1.1 Beschäftigten-Historik (BeH)

- Geringfügige Datumskorrekturen wurden durchgeführt.
- Ab 1999 sind nur die Personengruppen 101, 102, 103, 105, 106, 112, 118, 119, 120, 140, 141, 142, 143, 149, 201, 203 und 205 enthalten sowie die beiden Gruppen 109 und 209, die geringfügig Beschäftigte kennzeichnen. Nicht enthalten sind ab 1999 unter anderem kurzfristig Beschäftigte, d.h. Personengruppen 110, 202 und 210.
- Bis 1998 sind zusätzlich zu den oben aufgeführten Gruppen Meldungen mit fehlender Angabe der Personengruppe enthalten.
- Anmeldungen von sozialversicherungspflichtiger Beschäftigung werden herausgefiltert, da die Information über die Aufnahme von Beschäftigungsverhältnissen auch in den Jahresmeldungen vorhanden ist.

- Stornierungen werden berücksichtigt und die entsprechenden Sätze werden gelöscht. Änderungsmeldungen werden eingearbeitet.
- Duplikate werden nach folgenden Kriterien entfernt: Versicherungsnummer, Betriebsnummer, Beginn- und Enddatum, Abgabegrund, Entgelt und Erstelldatum.
- Sätze mit einem Entgelt in Höhe von Null bzw. mit fehlendem Entgelt sowie der Ausprägung 101 beim Merkmal Personengruppe und dem Wert 50 beim Abgabegrund (Jahresmeldung) werden nicht in die IEB übernommen.
- Versicherungsnummern werden zum Stand vom 30.06.2008 aktualisiert, stillgelegte werden gelöscht.
- Geschlecht und Geburtsdatum werden aus der Versicherungsnummer übernommen.
- Der Gebietsstand für Arbeitsort und Wohnort wird zum Stand vom 31.12.2008 aktualisiert.

3.1.2 Leistungsempfänger-Historik (LeH)

- Aus der LeH werden für die IEB nur Sätze für den Zeitraum ab 1.1.1975 (Enddatum) bis zum aktuellen Rand übernommen.
- Zu beachten ist, dass bei noch laufendem Bezug das Enddatum dem Ende des Bewilligungszeitraums entspricht und nicht dem Ende des Bezugszeitraums. Sätze mit leerem oder ungültigem Enddatum werden ausgeschlossen, da hier nicht davon ausgegangen werden kann, dass überhaupt eine Zahlung geleistet worden ist.
- Sätze ohne gültiges Beginndatum werden ausgeschlossen.
- Sätze mit einem Enddatum, das vor dem Beginndatum liegt, werden ausgeschlossen.
- Liegt das Enddatum für die Arbeitslosenhilfe einen Tag vor dem Beginndatum, so wird das Enddatum um ein Jahr verschoben.
- Sich überschneidende LeH-Sätze werden für die IEB entsprechend der folgenden Regeln bereinigt:
  - Alle Überschneidungen, bei denen in beiden Sätzen dieselbe Leistungsart und dieselbe Leistungshöhe steht, werden wie folgt bereinigt:
    - Wenn ein Satz vollständig innerhalb des Zeitraums eines anderen Satzes liegt, dann wird der Satz mit der kürzeren Dauer gelöscht. Bei gleicher Länge wird ein (zufällig ausgewählter) Satz gelöscht.

        - Bei Teilüberschneidungen wird das Enddatum des früher beginnenden Satzes korrigiert.
    - Überschneidungen zwischen Sätzen mit **unterschiedlicher Leistungsart** und/oder unterschiedlicher Leistungshöhe werden nur dann bereinigt, wenn es sich um eine Überschneidung von genau einem Tag handelt und die Originaldauer beider Sätze größer als ein Tag ist. In diesen Fällen wird das Enddatum des früher beginnenden Satzes um einen Tag zurückgesetzt.
- Der Gebietsstand von LeH Datensätzen ist auf den Stand vom 31.12.2008 bereinigt

3.1.3 Leistungs-Historik-Grundsicherung (LHG/XLHG)

- Es werden nur Sätze erwerbsfähiger Personen und von Personen unter 64 Jahren eingelesen.

3.1.4 Arbeitsuchenden-Historik (ASU/XASU)

- Die aus ASU verwendeten Daten umfassen den Zeitraum von 01/2000 bis 09/2009. Die registrierten Zeiten der Arbeitsuche der einzelnen Personen werden als vollständig angesehen.
- In der Datenbasis ASU gibt es nur die Unterscheidung zwischen Sätzen mit Status "arbeitslos" und "arbeitsuchend". In der IEB wird jedoch zwischen "arbeitslos", "krank" und "arbeitsuchend" 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 "arbeitsuchend" hat.
- In der XASU werden im Gegensatz zur ASU Krankheitsperioden nicht berücksichtigt, da in den XASU Daten keine Informationen zu Krankheiten vorliegen.
- Sätze mit einem Enddatum, das kleiner als das Beginndatum ist, werden ausgefiltert.
- Eine Gebietsstandsbereinigung wird nicht durchgeführt.

3.1.5 Bereinigungen der Gesamtdatei

Nachdem die Rohdaten bereinigt und zur Grundgesamtheit der IEB zusammengefasst wurden, erfolgt eine Bereinigung der Gesamtdatei der IEB über alle Quellen hinweg. Dabei werden folgende Bereinigungsschritte durchgeführt:

- Datensätze, bei denen das Alter kleiner 13 oder größer 75 ist, werden gelöscht.
- Datensätze, bei denen das Enddatum kleiner als das Beginndatum ist, werden gelöscht.
- Personenkonten mit einem Wechsel von Geschlecht oder Geburtsdatum werden nach Prioritätsregeln bereinigt.

3.2 Episodensplitting

Die SIAB-Daten liegen „gesplittet“ vor. Falls sich innerhalb eines Kontos Sätze überlappen, werden diese durch künstliche Sätze mit neuen Datumsangaben derart ersetzt, dass vollständig parallele bzw. überschneidungsfreie Zeiträume entstehen (Episodensplitting). Die Anzahl der Sätze erhöht sich dadurch.

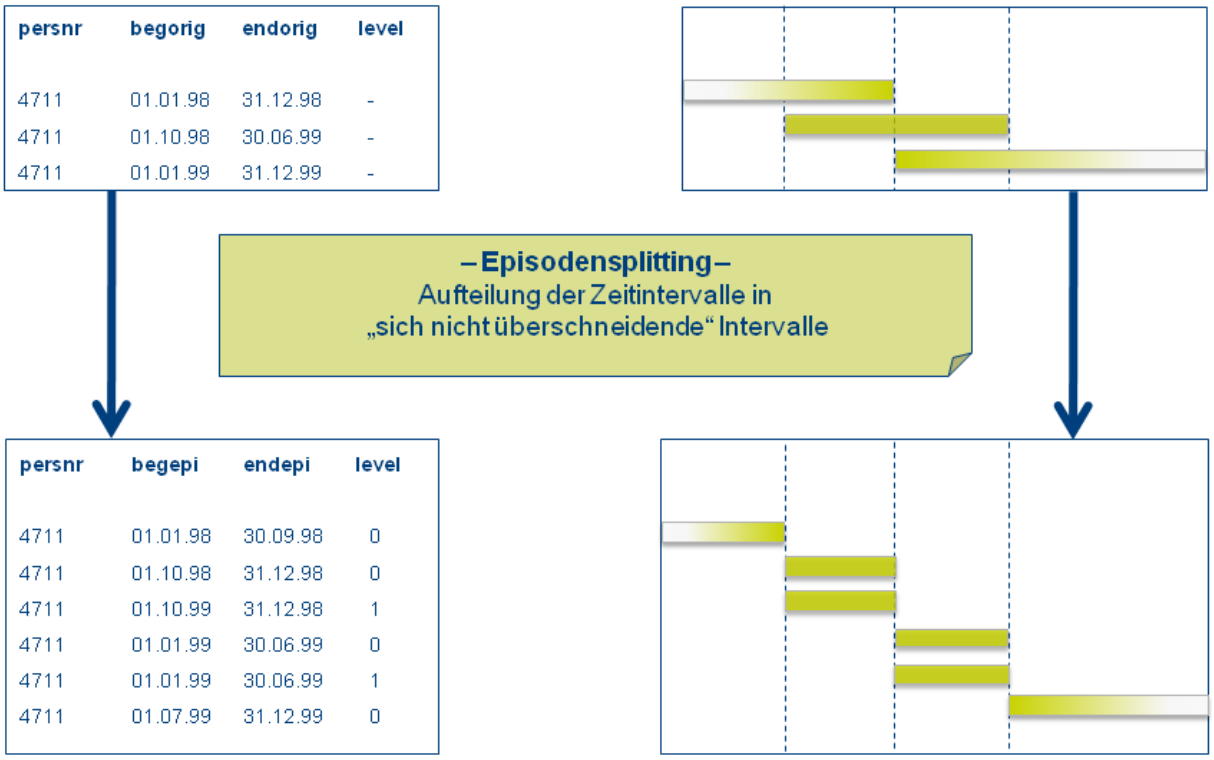

Abbildung 2: Episodensplitting

Durch das Episodensplitting werden Sätze verdoppelt und der Gültigkeitszeitraum geändert. Die ursprünglichen Datumsvariablen Beginn und Ende des Originalsatzes (begorig und endorig) bleiben erhalten, die Variablen Beginn und Ende der Episode (begepi und endepi)

kennzeichnen Anfang und Ende der gesplitteten Episoden. Ob Sätze gesplittet wurden, kann an einem Vergleich des ursprünglichen Zeitraums (begorig und endorig) mit dem Episodenzeitraum (begepi und endepi) festgestellt werden.

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

3.3 Stichprobenziehung

Die SIAB ist eine 2%-Stichprobe der Personen, für die ein Eintrag in den IEB vorliegt. Entsprechend lassen sich auf Basis der Stichprobe mit Hochrechnungsfaktor 50 Fallzahlen der Grundgesamtheit schätzen. Dies gilt grundsätzlich für die Personenkreise aller Datenquellen, allerdings ist der unterschiedliche Füllgrad der Quellen zu beachten. So wird beispielsweise das Bewerberangebot (ASU bzw. XASU) erst ab 2000 vollständig erfasst, davor nur zum Teil. Aus den ASU bzw. XASU-Sätzen vor 2000 kann also nicht auf die Grundgesamtheit der Bewerber geschlossen werden.

Somit ist die Ziehung der SIAB-Beobachtungen vergleichbar mit der Ziehung der IEBS.[3] Hinsichtlich der IABS gibt es jedoch zwei Unterschiede in der Stichprobenziehung. Zum einen ist die IABS eine Beschäftigtenstichprobe, d.h. die Personen wurden aus den Beschäftigten gezogen, LeH-Sätze wurden lediglich zugespielt. Es gibt in der IABS also keine Personen mit ausschließlich LeH-Sätzen, während es in der SIAB sehr wohl Personen gibt, die ausschließlich Nicht-BeH-Sätze aufweisen. Dadurch ist die SIAB repräsentativ für die Personenkreise aller Datenquellen, nicht nur für die Beschäftigten. Zum anderen wird die SIAB, im Gegensatz zur IABS, nicht nach Ost/West und deutsch/nichtdeutsch geschichtet gezogen. Aufgrund der hohen Fallzahl sind diese vier Personenkreise jedoch auch ohne explizite Schichtung mit annähernd gleicher Auswahlwahrscheinlichkeit in den Daten enthalten.

3.4 Fehlende Werte

In der SIAB gilt die folgende Codierung von fehlenden Werten:

Bezeichnung	Wert	Beschreibung
Keine (valide) Angabe	.z	Nicht systematisch fehlende Ausprägungen eines Merkmals, d.h. das Merkmal ist prinzipiell für die Quelle gefüllt, jedoch ist für die betrachtete Ausprägung keine Angabe vorhanden bzw. keine Zuordnung möglich.
Systematisch nicht gefüllt	.n	Ein Merkmal ist grundsätzlich für eine Quelle (dunkelgraue Felder der Merkmalsübersicht im Abschnitt 1.5) oder für einen gewissen Zeitraum nicht gefüllt.

[3] Die Ziehung der IEBS erfolgt über eine zufällige Auswahl von acht Geburtstagen (vgl. Oberschachtsiek et al. 2009).

4 Datenqualität und Probleme

4.1 Beschäftigten-Historik (BeH)

- Aufgrund der Einführung des Meldeverfahrens in den neuen Bundesländern kann bei BeH-Sätzen erst ab 1993 von einer hinreichenden Vollständigkeit der Ost-Meldungen ausgegangen werden. Der Anstieg der Zahl der BeH-Sätze ab 1999 folgt aus der Einführung der Meldepflicht für geringfügig Beschäftigte ab dem 1. April 1999.
- Innerhalb des Meldeverfahrens ist eine gewisse Zeitverzögerung nicht vermeidbar. Zwar müssen Änderungen von Beschäftigungsverhältnissen sofort gemeldet und bestehende Beschäftigungsverhältnisse jährlich bis zum April des folgenden Jahres bestätigt werden, faktisch treffen Meldungen jedoch zum Teil um Jahre verspätet ein. Die Historik-Datei 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 2009 z. B. kann die 18-Monats-Datei für das Jahr 2007 erstellt werden). Mehr als 3 Jahre verspätete Meldungen werden im IAB nicht berücksichtigt, sodass eine 36-Monats-Datei per definitionem einen Füllgrad von 100 % aufweist.
- In der den SIAB-Daten zugrundeliegenden Version der IEB liegt der Füllgrad der BeH-Meldungen im Jahr 2005 letztmalig bei 100 %. Für die Jahre 2006 und 2007 wurden 18-Monatsdateien verwendet, die Meldungen für das Jahr 2008 stammen aus einer 12-Monats-Datei. Es ist daher davon auszugehen, dass die BeH-Meldungen für die Jahre 2006, 2007 und insbesondere 2008 in der SIAB leicht 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 der SIAB 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 deutlich an[4].
- Für einzelne Geburtsjahrgänge zeigt sich im Jahr 1975 eine geringere Fallzahl als erwartet. Das Jahr 1975 ist somit für Auswertungen problematisch[5].

[4] Vgl. Bender et al. 1996.
[5] Vgl. Bender et al. 1996.

- Bedingt durch die Einführung des neuen Meldeverfahrens 1999 sind nur die Personengruppen 101, 102, 103, 105, 106, 112, 118, 119, 120, 140, 141, 142, 143, 201, 203 und 205 enthalten, sowie die beiden Gruppen 109 und 209, die geringfügig Beschäftigte kennzeichnen. Nicht enthalten sind unter anderem kurzfristig Beschäftigte (Personengruppen 110, 202 und 210). Ausnahme: Vor 1999 sind zusätzlich Meldungen erfasst, die keine (gültige) Angabe zur Personengruppe aufweisen, was insbesondere für das Jahr 1991 von Bedeutung ist, weil dort viele Beschäftigte in Ostdeutschland so gemeldet wurden.
- In einigen wenigen Fällen kann es innerhalb einer Episode zu einer falschen Reihenfolge der BeH-Sätze kommen. Evtl. wird ein BeH Satz mit der Ausprägung „.z“ für das Merkmal „Erwerbsstatus“ nachranging ausgewiesen, obwohl dieser Satz das höchste Entgelt ausweist.

4.2 Leistungsempfänger-Historik (LeH)

- Grundsätzlich ist zu beachten, dass bei noch laufendem Bezug das Enddatum dem Ende des Bewilligungszeitraums entspricht und nicht dem Ende des Bezugszeitraums. LeH-Sätze, die über den 14.09.2009 hinausreichen, bilden somit den Bewilligungszeitraum ab und sind in den SIAB-Daten, wenn auch nur zu einem sehr geringen Anteil (kleiner 1%), vorhanden.
- Ein sehr geringer Anteil der LeH-Sätze beginnt bereits vor 1975. Dies erklärt sich dadurch, dass in der IEB alle LeH-Sätze enthalten sind, die 1975 oder später enden. Für die neuen Bundesländer werden die LeH-Sätze erst ab 1992 vollständig erfasst.
- Die Leistungsbezugsdaten wurden früher auf Magnetbändern gesichert. Aufgrund eines Schadens an den Magnetbändern sind die Leistungsbezugsdaten bis einschließlich des Jahres 1980 nur teilweise enthalten. Daher ist auch in den SIAB-Daten davon auszugehen, dass Informationen zum Leistungsbezug im genannten Zeitraum nur zum Teil enthalten sind. Dies ist bei Analysen entsprechend zu berücksichtigen.

4.3 Leistungs-Historik-Grundsicherung (LHG/XLHG)

- Hinsichtlich der Vollständigkeit der Fallzahlen bzw. Leistungsbiografien aus den Quellen LHG und XLHG, bestehen in den IEB in den Jahren 2005 und 2006 deutliche Lücken. Ab 2007 existieren Lücken dann nur noch punktuell bei einzelnen Trägern. Es ist davon auszugehen, dass sich diese Lücken auch in den SIAB-Daten fortsetzen. Von einer Auswertung der Daten aus LHG und XLHG in den Jahren 2005 und 2006 wird daher dringend abgeraten. Entsprechende Analysen sollten frühestens ab dem Jahr 2007 beginnen. Darüber hinaus ist die Qualität der Meldungen in hohem Maß vom meldenden Träger abhängig, was besonders bei regionalen Analysen zu beach-

ten ist. Bei Auswertungen in eng definierten Querschnitten ist daher auf eine geeignete Trägerselektion zu achten.

- In den SIAB-Daten ist der Füllgrad des Merkmals „SGB-Einstellungsgrund“(„grund“, vgl. 5.4.1) relativ gering. Hier ist jedoch zu beachten, dass der „SGB-Einstellungsgrund“ ein rein personenbezogenes Merkmal ist. Wird daher ein Satz eines Individuums im ALG II Bezug aufgrund einer Änderung der Struktur seiner Bedarfsgemeinschaft gesplittet, weist der „SGB-Einstellungsgrund“ die Ausprägung „Missing“ (.z) auf.
- Beginn und Ende eines Sanktionssachverhalts führen zwar zu einem Splitting der Sätze, jedoch kann anhand der Daten weder die Dauer, noch die Art oder der Zeitpunkt der Verhängung bzw. des Beginns der Sanktion festgestellt werden.
- In der LHG als auch im höheren Maß in der XLHG ist zu beobachten, dass es überdurchschnittlich viele 15 und im geringeren Maß 16- und 17-Jährige gibt, die als Nicht-Erwerbsfähig eingestuft wurden. 15-, 16 bzw. 17-jährige Hilfebedürftige der zugelassenen kommunalen Träger sind deshalb möglicherweise schlechter in den IEB repräsentiert, da „Nicht Erwerbsfähige“ in die IEB nicht übernommen werden.
- In der Regel werden Wechsel von Trägerart oder Gebietszerschneidungen etwas später in den Daten abgebildet. Eine Ausnahme stellt das erste Halbjahr 2005 dar, da hier die geplante Gründung von ARGEn bereits in den Daten abgebildet ist. Der juristische Trägerformwechsel erfolgte aber erst etwas später.

4.4 Arbeitsuchenden-Historik (ASU/XASU)

4.4.1 ASU

- Die registrierten Zeiten der Arbeitsuche ab dem Endjahr 2000 für die einzelnen Personen werden als vollständig angesehen.
- Ältere Sätze sind nur für Arbeitsuchende enthalten, die ab 2000 oder vorher in den BA Systemen identifiziert werden konnten. Sätze mit Beginndaten vor 1990 oder Enddaten nach September 2009 dürften jedoch vorwiegend Eingabefehler sein.
- Bedingt durch Probleme während der Datenaufbereitung in den BA- bzw. IAB-Verfahren ist bei ASU-Sätzen eine eindeutige Zuordnung zu einer der Trägerarten der Rechtskreise S.GB III (BA) oder SGB II (ARGE, zugelassene kommunale Träger, getrennte Trägerschaft) nicht möglich. Das Merkmal „SGB II-Träger“ ist daher für alle ASU-Sätze statistisch nicht auswertbar und wurde daher für alle ASU-Sätze auf „Missing“ (.z) gesetzt.
- Von Mitte 2005 bis Mitte 2006 wurde in der BA das Verfahren coArb, aus dem die Arbeitsuchenden- und Bewerber-Angebotsdaten stammen, durch das Verfahren VerBIS

abgelöst. Ab Juli 2005 wurde das operative System 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. Betroffen von der VerBIS Umstellung sind die für die Quelle ASU gefüllten Ausprägungen der folgenden SIAB Merkmale:

  - Trägerart
  - Schwerbehindertenstatus
  - Staatsangehörigkeit
  - Ausbildung
  - Berufsordnung

- Eine Gebietsstandsbereinigung wird nicht durchgeführt.

4.4.2 XASU

- Im Gegensatz zu den Arbeitsuchenden-Meldungen aus den ARGEn und den geteilten Trägerschaften, zeigen sich seit 01.01.2005 für die 69 zugelassenen kommunalen Träger systematische Untererfassungen. Grundsätzlich ist erst ab Juni 2006 von einer Vollständigkeit der XASU-Sätze in der IEB und damit auch in den SIAB-Daten auszugehen.
- Wegen der fehlenden zentralen Zuordnung von eindeutigen Identifikationsnummern sind Brüche oder unplausible Arbeitsuche-Biografien in den IEB und damit auch in den SIAB-Daten möglich.
- Verschiedenste Merkmale in der SIAB weisen für die XASU zum Teil nur einen sehr geringen Füllgrad auf. Hiervon betroffen sind unter anderem die Merkmale Schulabschluss, Ausbildung, Schwerbehindertenstatus, Staatsangehörigkeit, Abgabegrund, Erwerbsstatus vor Arbeitsuche und die Angaben zum Wohnort. Die Füllgrade dieser Merkmale verbessern sich zum Teil im Zeitablauf. Das Merkmal Berufsordnung ist nahezu für den gesamten zur Verfügung stehenden Zeitraum in der XASU nicht ge-

füllt. Generell ist zu beachten, dass die Füllgrade zum Teil auch trägerspezifisch variieren.

- Weiterhin ist unklar, ob die Träger - wie vorgesehen - nur Zeiten melden, in denen sie selbst die Arbeitsuchenden betreuen, oder ob vorausgehende Zeiten der Betreuung durch die BA nach dem SGB III als durchgängige Episoden gemeldet werden, sofern der Arbeitsuche-Status nicht wechselt.
- Bei etlichen Trägern ist in den IEB durchgängig oder zeitweise der Anteil der gemeldeten Arbeitsuchenden an den gemeldeten erwerbsfähigen Hilfebedürftigen unplausibel hoch. Ein Grund dafür könnte eine möglicherweise fehlerhafte Bestimmung des Status „Nichtarbeitslos arbeitsuchend“ bei diesen Trägern sein.
- Die träger- und zeitraumabhängige Plausibilität der XASU-Angaben sollte jeweils unter Berücksichtigung der Untersuchungsfrage vor der Verwendung vom Nutzer geprüft werden.

5 Variablenbeschreibung

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

5.1 Identifikatoren

5.1.1 Systemfreie Personennummer (persnr)

Variablenlabel	Systemfreie Personennummer
Variablenname	persnr
Kategorie	Identifikatoren
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	Die systemfreie Personennummer zeigt an, welche Sätze zur gleichen Person gehören. Systemfrei bedeutet, dass von dieser Personennummer nicht auf Eigenschaften der Person oder Original-Identifikatoren geschlossen werden kann. Da es keinen einheitlichen BA-weiten Personen-Identifikator gibt, ist die Zuordnung von (BA-)Daten aus unterschiedlichen Datenquellen (z. B. Beschäftigung und Leistung) zu Personen nicht immer eindeutig. Die Bildung des quellenübergreifenden Personen-Identifikators basiert auf einer von der BA entwickelten Heuristik.

5.1.2 Systemfreie Betriebsnummer (betnr)

Variablenlabel	Systemfreie Betriebsnummer
Variablenname	betnr
Kategorie	Identifikatoren

Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Die systemfreie Betriebsnummer zeigt an, welche Sätze zum gleichen Betrieb gehören. Sie basiert auf der von der BA vergebenen Betriebsnummer, die durch eine systemfreie Nummer ersetzt wurde. Weitere Informationen zur Betriebsnummernvergabe der BA finden sich in Bender et al. 1996: S. 15 f. und S. 27-30. Dieses Merkmal ist mit einem anderen Variablenlabel auch in der Betriebsdatei enthalten. in der Betriebsdatei lautet das Variablenlabel: btr-id bhp 7508
Qualitätshinweis	Die Betriebsnummer fehlt in der SIAB in 9.080 (0,042%) BeH-Meldungen. Bei 6.522 (79%) dieser Meldungen handelt es sich um Meldungen der Personengruppe „205“ (Entgeltmeldungen für unständig Beschäftigte). Da Betriebsmerkmale (Arbeitsort, Wirtschaftszweig, Betriebsgröße usw.) über die originale Betriebsnummer zugespielt werden, fehlen in diesen Meldungen die Betriebsinformationen.

5.2 Gültigkeitszeitraum

5.2.1 Beginndatum Originalsatz (begorig)

Variablenlabel	Beginndatum Originalsatz
Variablenname	begorig
Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Das Beginndatum des Originalsatzes entspricht dem ursprünglichen Beginndatum der Meldung. Dies kann vom Beginndatum der Episoden (begepi) abweichen (vgl. S. 24). In BeH-Meldungen sind Beginn- und Endejahr wegen der Regeln des Meldeverfahrens immer identisch (Pflicht des Arbeitgebers zur Abgabe von Jahresmeldungen).

5.2.2 Enddatum Originalsatz (endorig)

Variablenlabel	Enddatum Originalsatz
Variablenname	endorig
Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Datum
Hierarchie	Keine

Detailbeschreibung	Das Enddatum des Originalsatzes entspricht dem ursprünglichen Enddatum der Meldung. Dies kann vom Enddatum der betreffenden Datenzeile, dem so genannten Episoden-Enddatum, abweichen. Grund dafür ist das so genannte Episodensplitting, vgl. S. 24. In BeH-Meldungen sind Beginn- und Endejahr wegen der Regeln des Meldeverfahrens immer identisch (Pflicht des Arbeitgebers zur Abgabe von Jahresmeldungen).

5.2.3 Beginndatum der Episode (begepi)

Variablenlabel	Beginndatum der Episode
Variablenname	begepi
Kategorie	Generierter Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Beginndatum der Episode. Mit dem Begriff „Episode“ wird der Zeitraum eines Satzes bezeichnet. Dieser Zeitraum kann vom Originalzeitraum der ursprünglichen Meldung abweichen (vgl. S. 24).

5.2.4 Enddatum der Episode (endepi)

Variablenlabel	Enddatum der Episode
Variablenname	endepi
Kategorie	Generierter Gültigkeitszeitraum
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Enddatum der Episode. Mit dem Begriff „Episode“ wird der Zeitraum eines Satzes bezeichnet. Dieser Zeitraum kann vom Originalzeitraum der ursprünglichen Meldung abweichen (vgl. S. 24).

5.3 Informationen zur Person

5.3.1 Geschlecht (frau)

Variablenlabel	Geschlecht
Variablenname	frau
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Geschlechter-Dummy (0 - Mann, 1 - Frau). Die Geschlechtsangabe ist pro Konto konstant.
Qualitätshinweis	In den Originaldaten kann es vorkommen, dass das Geschlecht wechselt. Dies wird bei der Datenaufbereitung bereinigt. Hierbei kommt Angaben aus Datenquellen der Leistungsgewährung (LeH, (X)LHG) die höchste Priorität zu.

5.3.2 Geburtsjahr (gebjahr)

Variablenlabel	Geburtsjahr
Variablenname	gebjahr
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Geburtsjahr ist pro Konto konstant.
Qualitätshinweis	In den Originaldaten kann es vorkommen, dass das Geburtsjahr zwischen den Quellen wechselt. Dies wird bei der Datenaufbereitung bereinigt. Hierbei kommt der Angabe aus der Sozialversicherungsnummer die höchste Priorität zu.

5.3.3 Staatsangehörigkeit (nation)

Variablenlabel	Staatsangehörigkeit
Variablenname	nation
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal enthält die Staatenschlüssel des Statistischen Bundesamtes. Vgl. 5.3.4.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt.

5.3.4 Staatsangehörigkeit vergröbert (nation_gr)

Variablenlabel	Staatsangehörigkeit vergröbert
Variablenname	nation_gr
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal fasst die Staatenschlüssel des Statistischen Bundesamtes zu Gruppen zusammen.

5.3.5 Familienstand (famst)

Variablenlabel	Familienstand
Variablenname	famst
Kategorie	Persönliche Merkmale
Herkunft	LeH, ASU
Datentyp	Numerisch
Hierarchie	Keine

Detailbeschreibung	Das Merkmal beschreibt den Familienstand. Der Familienstand schließt Kategorien der häuslichen Gemeinschaft mit Kindern oder Lebenspartnern ein. Allerdings ist nicht trennscharf, ob • eine verheiratete / in häuslicher Gemeinschaft lebende Person Kinder erzieht, • eine verheiratete / in häuslicher Gemeinschaft / alleinlebende Person überhaupt Kinder hat, • eine alleinerziehende / in häuslicher Gemeinschaft / alleinlebende Person verheiratet ist In der Quelle LeH umfasst das Merkmal nur zwei Ausprägungen, während in der Quelle ASU zwischen vier Ausprägungen unterschieden wird. Ein Abgleich der Informationen aus beiden Quellen wurde nicht vorgenommen.

5.3.6 Kinderzahl (kind)

Variablenlabel	Kinderzahl
Variablenname	kind
Kategorie	Persönliche Merkmale
Herkunft	LeH, ASU
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	Je nach Datenquelle hat das Merkmal eine unterschiedliche Bedeutung. Angegeben sind die Kinder im Haushalt unter 16 Jahren zum Zeitpunkt der Beantragung des Leistungsbezugs. **1) LeH** In der LeH wird nur unterschieden, ob die Anzahl an Kindern gleich Null oder mindestens Eins ist. Folgende Ausprägungen kommen daher vor: 0 Kein Kind 100 Mindestens ein Kind **2) ASU** Hier entspricht die Ausprägung des Merkmals der Anzahl an Kindern. Das Merkmal wird bei Änderungen der Leistungsart oder Bewilligungen nicht aktualisiert, sondern z.B. erst, wenn nach einer Beschäftigung ein erneuter Leistungsfall auftritt.

5.3.7 Ausbildung (bild)

Variablenlabel	Ausbildung
Variablenname	bild
Kategorie	Persönliche Merkmale
Herkunft	BeH, ASU, XASU
Datentyp	Numerisch
Hierarchie	keine

Detailbeschreibung	Zu beachten ist, dass das Merkmal je nach Quelle eine unterschiedliche Bedeutung hat **1) BeH** Das Merkmal enthält für BeH-Sätze die von den Arbeitgebern im Rahmen des Meldeverfahrens anzugebende Ausbildung. Bei dem Signierschlüssel „Ausbildung“ handelt es sich um einen kombinierten Schlüssel, durch den sowohl die erreichte Schulbildung als auch die abgeschlossene Berufsausbildung (jeweils höchster Abschluss) der/des Beschäftigten erhoben werden soll. Dabei ist grundsätzlich zunächst die Schulbildung festzustellen, danach die Berufsausbildung. Lediglich bei den Schlüsselpositionen 5 (Abschluss einer Fachhochschule) und 6 (Hochschul- / Universitätsabschluss) wird auf die Feststellung einer sonstigen Berufsausbildung verzichtet (vgl. BA 2005, S. IX.). Folgende Ausprägungen kommen vor: 1 Volks-/Hauptschule/mittlere Reife oder gleichwertige Schulbildung, ohne abgeschlossene Berufsausbildung 2 Volks-/Hauptschule/mittlere Reife oder gleichwertige Schulbildung, mit abgeschlossener Berufsausbildung 3 Mit Abitur, ohne abgeschlossene Berufsausbildung 4 Mit Abitur und abgeschlossener Berufsausbildung 5 Fachhochschulabschluss 6 Hochschulabschluss Aus den Daten kann nicht geschlossen werden, ob eine abgeschlossene Schulausbildung vorliegt oder nicht, beide Fälle können in den Ausprägungen „1“ und „2“ vorkommen. **2) ASU/XASU** Für ASU- und XASU-Sätze ist die abgeschlossene Berufsausbildung zum Beginn der Arbeitsuche enthalten. Folgende Ausprägungen kommen vor: 21 ohne abgeschlossene Berufsausbildung 22 betriebliche Ausbildung 23 außerbetriebliche Ausbildung 24 Berufsfachschule 25 Fachschule 26 Fachhochschule 27 Universität „Zusammen mit einem Betriebswechsel treten häufig auch Änderungen im Ausbildungsstatus auf. Das liegt daran, dass beim neuen Betrieb die Meldedaten neu zusammenstellt 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) Eine Methode, fehlende Werte oder inkonsistente zeitliche Verläufe der Bildungsangaben in der Vorgängerstichprobe IABS zu bereinigen, fin-

	det 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.
Qualitätshinweis	„Für bestimmte Teilgruppen gibt es einen hohen Anteil fehlender Angaben, weil das Merkmal versicherungsrechtlich keine besondere Bedeutung hat (das betrifft z.B. die geringfügig Beschäftigen).“ (Meinken/Koch 2004, S. 63).

5.3.8 Schulausbildung (schbild)

Variablenlabel	Schulausbildung
Variablenname	schbild
Kategorie	Persönliche Merkmale
Herkunft	ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal enthält den Schulabschluss zu Beginn der Arbeitslosigkeit bzw. Arbeitsuche.
Qualitätshinweis	Der Füllgrad in der XASU steigert sich zwar über die Jahre, bleibt aber relativ gering (zum Teil unter 50%).

5.3.9 Qualifikation (quali)

Variablenlabel	Qualifikation
Variablenname	Quali
Kategorie	Persönliche Merkmale
Herkunft	ASU
Datentyp	Numerisch
Hierarchie	Keine

Detailbeschreibung	Bei diesem Merkmal wird nicht nur die Ausbildung berücksichtigt, sondern auch die Berufserfahrung. Bei gleichwertiger Befähigung aufgrund von Vorbildung, Leistung oder Erfahrung kann auch ohne einen entsprechenden Abschluss eine Qualifikationsstufe eingetragen werden.

5.3.10 Schwerbehindertenstatus (schweb)

Variablenlabel	Schwerbehindertenstatus
Variablenname	Schweb
Kategorie	Persönliche Merkmale
Herkunft	ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Die Anerkennung des Schwerbehindertenstatus erfolgt bei einem korrespondierenden Grad der Behinderung von über 50 %. Die Gleichstellung erfolgt nach §2 SchwbG auf Antrag bei einem korrespondierenden Grad der Behinderung von 30 bis 50 %. Das Merkmal bezieht sich auf das Beginndatum Arbeitslosigkeit bzw. des Arbeitsuchendenstatus.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt.
Qualitätshinweis	Der Füllgrad des Merkmals ist in der XASU in den Jahren 2005 und 2006 schlecht.

5.3.11 Gesundheitliche Einschränkungen (gesund_ein)

Variablenlabel	Gesundheitliche Einschränkungen
Variablenname	gesund_ein
Kategorie	Persönliche Merkmale
Herkunft	ASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Die Variable informiert darüber, ob Einschränkungen vorliegen.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt.
Qualitätshinweis	Diese Variable ist von der VerBIS-Umstellung betroffen (siehe Abschnitt 4.4.1). Die Datenqualität des Merkmals ist als gut zu bezeichnen.

5.4 Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

5.4.1 Abgabegrund/ Beendigungsgrund/ SGB-II-Einstellungsgrund /Abmeldegrund (grund)

Variablenlabel	Abgabegrund /Beendigungsgrund/ SGB-II-Einstellungsgrund /Abmeldegrund
Variablenname	Grund

Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	**1) BeH** In BeH-Meldungen kennzeichnet der Abgabegrund den Anlass, aus dem der Arbeitgeber den Sozialversicherungsträgern die betreffende Beschäftigungsmeldung übermittelt hat. Allerdings treten in der IEB und damit auch in der SIAB nicht alle im Rahmen des Meldeverfahrens möglichen Abgabegründe auf. So beinhaltet die IEB nur Meldungen mit Entgeltangaben (also Jahres-, Unterbrechungs- und Abmeldungen), während Anmeldungen wegen fehlender Entgeltinformationen nicht enthalten sind. Ein Informationsverlust ist damit jedoch nicht verbunden, da die Angaben aus einer Anmeldung mit der darauf folgenden Jahres-, Unterbrechungs- oder Abmeldung erneut übermittelt werden. Die Abgabegründe für Beschäftigungsmeldungen sind in der SIAB nach den Regeln des seit 01.01.1999 geltenden Meldeverfahrens (gemäß DEÜV) kodiert. **2) LeH** In LeH-Meldungen gibt das Merkmal den Grund für das Ende des Bezugs von Arbeitslosengeld, -hilfe oder Unterhaltsgeld an. Über die Gründe für den Beginn des Leistungsbezugs finden sich in der LeH und damit auch in der SIAB hingegen keine Informationen, da die LeH mit den Meldungen der Arbeitsagenturen an die Krankenkassen über abgeschlossene Leistungsbezugsdauern gefüllt wird. **3) LHG/XLHG** Bei LHG- und XLHG-Datensätzen ist der SGB-II-Einstellungsgrund enthalten und gibt den Grund für die Beendigung von laufenden Leistungen an. Der SGBII-Beendigungsgrund **bezieht sich auf die Person**, nicht aber auf die Bedarfsgemeinschaft. Das Merkmal ist genau zum Satz-Ende gültig. Bei einem Wechsel seines Wertes wird der Datensatz gesplittet. Es ist zu beachten, dass der SGB-II-Beendigungs-/Einstellungsgrund ein personenbezogenes Merkmal ist. Wird bei einem anderen als dem betrachteten Mitglied der Bedarfsgemeinschaft der SGBII-Leistungsbezug beendet, werden alle Sätze der Mitglieder in einer Bedarfsgemeinschaft an diesem Datum gesplittet, der SGBII-Beendigungs-/Einstellungsgrund ist jedoch nur für die betroffene Person gefüllt. **4) ASU/XASU** Bei ASU-Sätzen ist der Abmelde- bzw. Abgangsgrund enthalten. Hierbei ist zu beachten, dass ab dem 26.04.2003 eine Reduktion der Merkmalsausprägungen erfolgte. Die alten Werte sind, soweit es möglich war, auf die neuen Werte umgeschlüsselt.

Qualitätshinweis	Anteil valider Werte (Füllgrad) des Abgabegrundes in den Quellen LHG, XLHG, ASU und XASU:
	`| 2005 2006 2007 2008 2009`
	`----------+----------`
	`LHG | 0.168 0.109 0.169 0.135 0.057`
	`XLHG | 0.015 0.065 0.093 0.108 0.133`

	ASU | 0.786 0.926 0.943 0.889 0.555 XASU | 0.247 0.509 0.491 0.445 0.267

5.4.2 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	**1) BeH** In BeH-Sätzen weist das Merkmal jeweils das Brutto-Tagesentgelt des Beschäftigten aus. Es wurde berechnet aus dem vom Arbeitgeber gemeldeten Zeitraumentgelt sowie der Dauer des Zeitraums der ungesplitteten Originalmeldung in Kalendertagen. Das Tagesentgelt ist für den gesamten Zeitraum in Euro angegeben. 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, wurden nur bis zu dieser Grenze gemeldet. Hierbei ist zu beachten, dass es in der gesetzlichen Rentenversicherung zwei Beitragsbemessungsgrenzen gibt: die Grenze der Arbeiterrenten- und Angestelltenversicherung und die Grenze der Knappschaftlichen Rentenversicherung, die meist höher liegt. Seit der Einbeziehung der geringfügig Beschäftigten in das Meldeverfahren zum 01.04.1999 werden auch Entgelte unterhalb der Geringfügigkeitsgrenze erfasst; die Beitragsbemessungsgrenze als obere Kappungsgrenze gilt jedoch weiterhin. In einigen Fällen übersteigen die gemeldeten Entgelte die Beitragsbemessungsgrenze allerdings trotzdem. Dies dürfte in der Regel auf Jahressonderzahlungen zurückzuführen sein, die der Arbeitgeber bei den Jahres-, Unterbrechungs- oder Abmeldungen zum normalen Entgelt hinzurechnen kann. Dabei ist es dann unerheblich, ob durch diese Addition die für den Meldezeitraum maßgebliche Beitragsbemessungsgrenze in der Rentenversicherung überschritten wird. Es könnte sich allerdings auch um fehlerhafte Angaben bei Beschäftigungszeit oder Entgelthöhe handeln, was wegen der Versicherungsrelevanz dieser Informationen jedoch eher selten der Fall sein dürfte. 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. 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. 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, also Meldungen

	mit (Zeitraum-)Entgelt = 0, mit der Bedingung Tagesentgelt = 0 identifziert werden. **2) LeH** Bei LeH-Datensätzen enthält das Merkmal den täglichen Leistungssatz, jeweils umgerechnet in Euro. Es ist zu berücksichtigen, dass der tägliche Leistungssatz bei Datensätzen mit einem Original-Beginndatum, das vor dem 01.01.1998 liegt, für Werktage und bei Datensätzen mit einem Original-Beginndatum ab 01.01.1998 für Kalendertage gilt. Die Angabe von 0 Euro ist seit 01.01.2005 auf Sperrzeiten oder Leistungsunterbrechungen zurückzuführen. Ist für einen Satz mit Leistungssatz gleich 0 ein Abmeldegrund/Beendigungsgrund /Abgabegrund angegeben, so handelt es sich um eine Unterbrechungsmeldung. Bei Sätzen, die eine Sperrzeit widerspiegeln, ist der Anspruch derselbe wie vor Beginn der Sperrzeit.

5.4.3 Gleitzone (gleitz)

Variablenlabel	Gleitzone
Variablenname	gleitz
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal ist erst ab 2003 und nur für BeH-Datensätze gefüllt. Es zeigt an, ob es sich bei der Beschäftigungsmeldung um eine Beschäftigung im Niedriglohnbereich innerhalb der sogenannten Gleitzone mit einem monatlichen Bruttoarbeitsentgelt von 400,01 bis 800,00 € handelt (sog. Midi-Job), für die der Arbeitnehmer nur einen reduzierten Beitrag zum Gesamtsozialversicherungsbeitrag leisten muss. Da Arbeitnehmer mit einem Entgelt in der Gleitzone freiwillig den „normalen“ Sozialversicherungsbeitrag zahlen können, sind nicht automatisch alle Beschäftigungen mit einem entsprechen Entgelt als Gleitzone gekennzeichnet. Die entsprechende gesetzliche Regelung besteht seit dem 01.04.2003.

5.4.4 Beruf – ausgeübte Tätigkeit (beruf)

Variablenlabel	Beruf – ausgeübte Tätigkeit
Variablenname	beruf
Kategorie	Beschäftigung/ Leistungsbezug
Herkunft	BeH, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine

Detailbeschreibung	**1) BeH** Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“. Der Arbeitgeber verschlüsselt die Tätigkeit des Beschäftigten hierbei gemäß der „Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen“ (Hrsg. von der Bundesanstalt für Arbeit, Nürnberg 1988), in dem ca. 25.000 Berufsbezeichnungen zu finden sind. Die Berufsordnung besteht aus einer 3-stelligen Kennziffer und umfasst rund 330 Ausprägungen. 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. BA 2005, S. V). **2) ASU/XASU** Bei ASU- und XASU-Datensätzen bezeichnet das Merkmal den Beruf der letzten Tätigkeit. Das Merkmal bezieht sich auf den Stand zu Beginn des Arbeitsuche-Datensatzes.

5.4.5 Stellung im Beruf und Arbeitszeit (stib)

Variablenlabel	Stellung im Beruf und Arbeitszeit
Variablenname	stib
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Die berufliche Stellung des Beschäftigten im Meldezeitraum wird vom Arbeitgeber im Rahmen der „Angaben zur Tätigkeit“ übermittelt. Das Merkmal „Berufsstellung“ unterscheidet zunächst zwischen Vollzeit- und Teilzeitbeschäftigten; maßgeblich ist dabei das Verhältnis zwischen der arbeitsvertraglich vereinbarten und der betriebsüblichen Arbeitszeit. Nur für Vollzeitbeschäftigte liefert die Variable tatsächlich Angaben zur beruflichen Stellung, während für Teilzeitbeschäftigte nur erfasst wird, ob ihre Stundenzahl eine bestimmte Grenze übersteigt oder nicht. Diese Grenze lag bis 1978 bei 20 Wochenstunden, zwischen 1979 und 1987 bei 15 Wochenstunden und seit 1988 bei 18 Wochenstunden. Die Unterscheidung zwischen vollzeitbeschäftigten Arbeitern (1,2) und Angestellten richtet sich ausschließlich nach der Art der Rentenversicherungspflicht (BfA für Angestellte und LVA für Arbeiter). Neben den Auszubildenden, Praktikanten und Volontären zählen auch Anlernlinge, Schüler an Schulen des Gesundheitswesens und Teilnehmer an geförderten Maßnahmen zur beruflichen Fortbildung, Umschulung und betrieblichen Einarbeitung zur Kategorie Beschäftigte in beruflicher Ausbildung (0). Falls mehrere Schlüsselzahlen in Betracht kommen, ist der Arbeitgeber dazu angehalten, die Einstufung nach der überwiegend ausgeübten Tätigkeit vorzunehmen. Lässt sich diese nicht eindeutig feststellen, ist die Schlüsselzahl der höherwertigen Stellung im Beruf einzutragen (vgl. BA (2005): S. VI).

5.4.6 Erwerbsstatus (erwstat)

Variablenlabel	Erwerbsstatus
Variablenname	erwstat
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Merkmal nimmt für jede Datenquelle unterschiedliche Werte mit unterschiedlichen Bedeutungen an. **1) BeH** Für BeH-Sätze ist die Personengruppe der Beschäftigtenmeldung enthalten, die mit dem neuen Meldeverfahren zum 01.01.1999 eingeführt wurde. Sie weist beitrags- oder leistungsrechtliche Besonderheiten des Beschäftigungsverhältnisses aus, die mindestens einen Sozialversicherungsträger betreffen, unterschieden nach meldenden Institutionen. Sozialversicherungspflichtige Arbeitnehmer, deren Beschäftigungsverhältnis keine Besonderheiten aufweist, erhalten den Schlüssel 101. Treffen mehrere Schlüssel zu, ist der kleinste anzugeben. Geringfügig Beschäftigte lassen sich in der SIAB als die Personengruppen 109 und 209 abgrenzen. Das Merkmal Personengruppe ist erst seit der Einführung des neuen Meldeverfahrens zum 01.01.1999 Bestandteil der Meldungen zur Sozialversicherung. Die Personengruppe kann in Meldungen, 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. Die Meldepflicht für geringfügig Beschäftigte besteht seit 01.04.1999. Vorher sind keine Daten über geringfügig Beschäftigte enthalten. **2) LeH** Für LeH-Sätze ist die gruppierte Leistungsart angegeben, d.h. es ist angegeben, ob Arbeitslosengeld, Arbeitslosenhilfe, Unterhaltsgeld bezogen werden oder ob Beiträge zur privaten Pflegeversicherung von der BA übernommen werden. Die Ausprägungen dieses Merkmals wurden durch Zusammenfassung der Ausprägungen des Merkmals Leistungsart generiert. **3) ASU** Bei ASU ist der Status der Arbeitsuche enthalten. Als Ratsuchende werden Empfänger von Arbeitslosengeld (bzw. Arbeitslosengeld II), die Leistungen unter den erleichterten Voraussetzungen des §428 SGB III (bzw. des §65 Abs. 4 SGB II) beziehen, sowie Nichtleistungsempfänger ohne Vermittlungsbereitschaft i. S. des § 252 Abs. 8 SGB VI, wenn das 58. Lebensjahr vollendet wurde, geführt. **4) XASU** In der XASU werden bislang nur Episoden mit dem Status • „nicht arbeitslos, aber arbeitsuchend“

	• „arbeitsuchend und gleichzeitig arbeitslos“ gemeldet bzw. berechnet.

5.4.7 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	Die Variable zeigt für ASU und XASU-Sätze den Erwerbsstatus vor der Arbeitsuche an. Ab 12/2002 erfolgte eine deutliche Reduzierung der Ausprägungen. Die Werte älterer Sätze wurden auf die aktuell gültigen Werte umgeschlüsselt, die somit für den gesamten Zeitraum gültig sind. Da das Merkmal vor dem Episodensplitting gebildet wird, ist die Angabe für den ungesplitteten Originalssatz und für die erste gesplittete Episode gültig.

Qualitätshinweis	Der Anteil valider Werte (Füllgrad) ist für XASU-Sätze sehr gering:
	`| 2005 2006 2007 2008 2009`
	`----------+----------`
	`XASU | 0.014 0.057 0.158 0.232 0.410`

5.4.8 Leistungsart (lart)

Variablenlabel	Leistungsart
Variablenname	lart
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LeH
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	Art der von der BA gezahlten Entgeltersatzleistung. (Im Merkmal „Erwerbsstatus“ wird dieses Merkmal für die Quelle LeH zusammengefasst.)

5.4.9 BA-Kundengruppe (kundengr)

Variablenlabel	BA-Kundengruppe
Variablenname	kundengr
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU
Datentyp	numerisch

Hierarchie	keine
Detailbeschreibung	Das Merkmal enthält die Eingruppierung der BA-Kunden in verschiedene Typen wie Marktkunde, Beratungskunde (aktivieren), Beratungskunde (fördern) sowie Betreuungskunde. Diese Unterscheidung berücksichtigt auch den Handlungsbedarf hinsichtlich potenziell vorhandender individueller Integrationshemmnisse.

5.4.10 Art der Kündigung (art_kuend)

Variablenlabel	Art der Kündigung der letzten Tätigkeit
Variablenname	art_kuend
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU
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.

5.4.11 Bereitschaft zur bundesweiten Stellensuche (mobil)

Variablenlabel	Bereitschaft zur bundesweiten Stellensuche
Variablenname	mobil
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	ASU
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	Dies Merkmal beschreibt die regionale Ausgleichsfähigkeit einer Person. Bei Meldungen zur Arbeitsuche muss die Person angeben, ob bundesweit oder regional eingeschränkt nach Arbeit gesucht wird. Wenn nur ausgewählte Regionen angegeben werden, erfolgt die Nachfrage, ob trotzdem auch andere Regionen in Betracht gezogen werden können. Wenn dies der Fall ist, gilt eine Person als "regional ausgleichsfähig", andernfalls nicht. Bei bundesweiter Suche gilt eine Person dementsprechend automatisch als "ausgleichsfähig".

5.4.12 Arbeitszeit des Stellengesuchs (arbzeit)

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

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

5.4.13 Restanspruchsdauer Arbeitslosengeld (restanspruch)

Variablenlabel	Restanspruch Arbeitslosengeld
Variablenname	restanspruch
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LeH
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	Der Restanspruch ist der Anspruch, der nach Ende des aktuellen Leistungsbezugs übrig bleibt. Der Restanspruch bleibt sieben Jahre bestehen. Liegt das Enddatum des Leistungsempfangs vor dem 01.01.1998, wird der Restanspruch in Werktagen angegeben. Danach erfolgt die Angabe in Kalendertagen. Dieser Anspruch zählt maximal bis auf die Höchstdauer für das jeweilige Lebensalter.

5.4.14 Trägerart (traeger)

Variablenlabel	Trägerart
Variablenname	traeger
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	LHG, XLHG, XASU
Datentyp	Numerisch
Hierarchie	keine
Detailbeschreibung	Das Merkmal zeigt die drei möglichen Trägerschaftsarten des Rechtskreises SGB II (Arbeitsgemeinschaften (ARGEn), optierende Kommunen, getrennte Träger) und die Trägerschaftsart des Rechtskreises SGB III (Bundesagentur für Arbeit). Arbeitsgemeinschaften (ARGEn) bilden sich aus einer gemeinsamen Aufgabenwahrnehmung von BA und Kommunen. Bei den optierenden Kommunen übernimmt die jeweilige Kommune die Aufgaben des SGB II in Eigenregie. Bei der geteilten Trägerschaft bleibt die Aufgabenteilung zwischen BA und Kreis bestehen. Zu beachten ist, dass die Grenzen der Träger in Bezug auf Kreise nicht trennscharf sind und somit nicht mit den Grenzen der Kreise übereinstimmen müssen. Bei XASU-Sätzen ist die Art des Trägers enthalten, der das Bewerberangebot führt. Die Information wird aus dem Bewerberangebot jeweils zu Beginn des ungesplitteten Satzes zugespielt.

5.5 Betriebsmerkmale

5.5.1 Wirtschaftszweig 73 – Wirtschaftsgruppe 73 (w73_3)

Variablenlabel	w73
Variablenname	w73_3
Kategorie	Betriebsmerkmale
Herkunft	BeH

Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt | Wirtschaftsabteilung | Wirtschaftsgruppe
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WS73 als 3-Steller aus. Die Wirtschaftsklasse W73 ist von 1975 bis einschließlich 2002 gefüllt. WS73 steht für das "Verzeichnis der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit, Ausgabe 1973“. Mittels eines dreistelligen Zahlencodes werden 269 Wirtschaftsklassen unterschieden, wobei die erste Stelle dieses Codes die Wirtschaftsabteilung, insgesamt 10, definiert und die beiden ersten Stellen zusammen die jeweilige Gruppe, insgesamt 95, festlegen. Jedem Betrieb darf nur eine Kennziffer zugeordnet werden. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, muss die wirtschaftliche Haupttätigkeit ermittelt werden.

5.5.2 Wirtschaftszweig 93 – Wirtschaftsunterklasse 93 (w93_5)

Variablenlabel	w93 unterklasse
Variablenname	w93_5
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt | Wirtschaftsabteilung | Wirtschaftsgruppe | Wirtschaftsklasse | Wirtschaftsunterklasse
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ93 als 5-Steller aus. WZ93 steht für die „Klassifikation der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit, Ausgabe 1993“. Der Wirtschaftszweig nach der 93er Klassifikation ist von 1999 bis 2003 enthalten. Jedem Betrieb darf nur eine Kennziffer zugeordnet werden. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, muss die wirtschaftliche Haupttätigkeit ermittelt werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller ausgewiesen.

5.5.3 Wirtschaftszweig 93 – Wirtschaftsgruppe 93 (w93_3)

Variablenlabel	w93 gruppe
Variablenname	w93_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt | Wirtschaftsabteilung | Wirtschaftsgruppe

Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ93 als 3-Steller aus. WZ93 steht für die „Klassifikation der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit, Ausgabe 1993“. Der Wirtschaftszweig nach der 93er Klassifikation ist von 1999 bis 2003 enthalten. Jedem Betrieb darf nur eine Kennziffer zugeordnet werden. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, muss die wirtschaftliche Haupttätigkeit ermittelt werden.

5.5.4 Wirtschaftszweig 03 – Wirtschaftsunterklasse 03 (w03_5)

Variablenlabel	w03 unterklasse
Variablenname	w03_5
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt | Wirtschaftsabteilung | Wirtschaftsgruppe | Wirtschaftsklasse | Wirtschaftsunterklasse
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ03 als 5-Steller aus. WZ03 steht für die „Klassifikation der Wirtschaftszweige Ausgabe 2003“ des Statistischen Bundesamtes (Hrsg.). Der Wirtschaftszweig nach der 03er Klassifikation ist erst seit 2003 gefüllt. Jedem Betrieb darf nur eine Kennziffer zugeordnet werden. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, muss die wirtschaftliche Haupttätigkeit ermittelt werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller ausgewiesen.

5.5.5 Wirtschaftszweig 03 – Wirtschaftsgruppe 03 (w03_3)

Variablenlabel	w03 gruppe
Variablenname	w03_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabteilung | Wirtschaftsgruppe | Wirtschaftsklasse
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ03 als 3-Steller aus. WZ03 steht für die „Klassifikation der Wirtschaftszweige Ausgabe 2003“ des Statistischen Bundesamtes (Hrsg.). Der Wirtschaftszweig nach der 03er Klassifikation ist erst seit 2003 gefüllt. Jedem Betrieb darf nur eine Kennziffer zugeordnet werden. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, muss die wirtschaftliche Haupttätigkeit ermittelt werden.

5.5.6 Wirtschaftszweig 08 – Wirtschaftsunterklasse 08 (w08_5)

Variablenlabel	w08 unterklasse
Variablenname	w08_5
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt | Wirtschaftsabteilung | Wirtschaftsgruppe | Wirtschaftsklasse | Wirtschaftsunterklasse
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ08 als 5-Steller aus. WZ08 steht für die „Klassifikation der Wirtschaftszweige Ausgabe 2008“ des Statistischen Bundesamtes (Hrsg.). Der Wirtschaftszweig nach der 03er Klassifikation ist erst seit 2008 gefüllt. Jedem Betrieb darf nur eine Kennziffer zugeordnet werden. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, muss die wirtschaftliche Haupttätigkeit ermittelt werden.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird der Wirtschaftszweig lediglich als 3-Steller ausgewiesen.

5.5.7 Wirtschaftszweig 08 – Wirtschaftsgruppe 08 (w08_3)

Variablenlabel	w08 gruppe
Variablenname	w08_3
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Wirtschaftsabschnitt | Wirtschaftsabteilung | Wirtschaftsgruppe
Detailbeschreibung	Dieses Merkmal weist den Wirtschaftszweig gemäß der Klassifikation WZ08 als 3-Steller aus. WZ08 steht für die „Klassifikation der Wirtschaftszweige Ausgabe 2008“ des Statistischen Bundesamtes (Hrsg.). Der Wirtschaftszweig nach der 08er Klassifikation ist erst seit 2008 gefüllt. Jedem Betrieb darf nur eine Kennziffer zugeordnet werden. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, muss die wirtschaftliche Haupttätigkeit ermittelt werden.

5.5.8 Erstes Auftreten des Betriebes (grd_jahr)

Variablenlabel	1. auftreten btr.
Variablenname	grd_jahr
Kategorie	Betriebsmerkmale
Herkunft	BeH
Hierachie	Keine

Detailbeschreibung	Bei dieser Variablen handelt es sich um das erste Auftreten der Nummer des Betriebes im Datensatz. Wird eine Betriebsnummer erst nach 1975 erstmalig nachgewiesen, könnte es sich um das Gründungsdatum des betreffenden Betriebes handeln. Allerdings könnte es sich auch um einen bereits länger existierenden Betrieb handeln, dem nach einem Wechsel des Inhabers oder einer Änderung der Rechtsform eine neue Betriebsnummer zugeordnet wurde. Hierbei verfahren die Arbeitsagenturen jedoch nicht einheitlich. Auch bei Ausgliederungen von Unternehmensteilen vergeben die Arbeitsagenturen zum Teil neue Betriebsnummern.

5.5.9 Letztes Auftreten des Betriebes (lzt_jahr)

Variablenlabel	Letzt. auftreten btr.
Variablenname	lzt_jahr
Kategorie	Betriebsmerkmale
Herkunft	BeH
Hierachie	Keine
Detailbeschreibung	Bei dieser Variablen handelt es sich um das letztmalige Auftreten der Nummer des Betriebes im Datensatz. Zur Betriebsnummernvergabe vgl. auch 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.).

5.5.10 Anzahl der Beschäftigten gesamt (az_ges)

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

5.5.11 Anzahl Vollzeitbeschäftigte (az_ges_vz)

Variablenlabel	anz. alle b. vz
Variablenname	az_ges_vz
Kategorie	Generierte Betriebsmerkmale
Herkunft	BeH
Hierachie	Keine
Detailbeschreibung	Die Variable enthält die Anzahl der Vollzeitbeschäftigten eines Betriebes. Nicht enthalten sind hier die geringfügig Beschäftigten sowie ruhende Beschäftigungsverhältnisse (Tagesentgelt von Null).

5.5.12 Anzahl geringfügig Beschäftigter (az_gf)

Variablenlabel	anz. gfg. b.
Variablenname	az_gf
Kategorie	Generierte Betriebsmerkmale
Herkunft	BeH
Hierachie	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 es erst seit diesem Jahr in das Meldeverfahren zur Sozialversicherung integriert wurde.

5.5.13 Betriebsentgelt (te_med)

Variablenlabel	median lohn alle vz-b.
Variablenname	te_med
Kategorie	Generierte Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Angegeben ist der Median der Tagesentgelte der sozialversicherungspflichtig Beschäftigten (ohne Teilzeitbeschäftigte, ohne geringfügig Beschäftigte, ohne Beschäftigte mit Entgelt gleich 0) des jeweiligen Betriebes zum 30.06. Das Betriebsentgelt ist immer in Euro angegeben.

5.6 Ortsangaben

5.6.1 Wohnort Kreis (wo_kreis)

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

Detailbeschreibung	In BeH- und LeH-Sätzen liegt der Kreis des Wohnortes erst für die Jahre ab 1999 vor. Darin ist jeweils der Kreis (kreisfreie Stadt oder Landkreis) angegeben, in dem der Versicherte wohnt. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des Bundeslandes, in der ersten bis dritten Stelle den Regierungsbezirk und in der ersten bis fünften Stelle den Kreis. Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null. Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben in der BeH und LeH auf den Gebietsstand vom 31.12.2008 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2008 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. Für andere Quellen als die BeH und LeH erfolgte keine Gebietsstandsaktualisierung. Für LHG-Sätze ist das Merkmal ab Januar 2005 gefüllt, für XLHG- und XASU-Sätze frühestens ab November 2008.
Besonderheiten	Dieses Merkmal wird Gastwissenschaftler/-innen wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt.

5.6.2 Wohnort Arbeitsagentur (wo_aa)

Variablenlabel	Wohnort Arbeitsagentur
Variablenname	wo_aa
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Regionaldirektion | Arbeitsagentur
Detailbeschreibung	Das Merkmal enthält die Arbeitsagentur des Wohnorts des BA-Kunden. Für die Quellen LHG und XLHG gilt der Wohnort für den Zeitraum des Satzes. Ein Wechsel des Wohnorts (bereits auf Ebene der Geschäftsstellen bzw. Gemeinden) löst einen Split des Satzes aus. Bei ASU und XASU enthält das Merkmal den Wohnort zum Beginn der Arbeitslosigkeit bzw. des Arbeitsuchendenstatus. Das Merkmal ist für die BeH und LeH ab 1999 vorhanden, für LHG, XLHG und XASU ab 2005. Für ASU Sätze ist das Merkmal bereits vor 1999 gefüllt. **1) BeH, LeH** In BeH-Sätzen wird der Arbeitsagenturbezirk des Wohnorts erst ab 1999 ausgewiesen. Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Arbeitsagenturangaben für BeH und LeH Sätze auf den Gebietsstand vom 31.12.2008 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Wohnortes zu einem Arbeitsamtsbezirk nach den Grenzen, welche die Arbeitsagenturbezirke zum 31.12.2008 hatten. Da sich die Grenzen der Arbeitsagenturbezirke im Zeitverlauf geändert haben, würden ohne diese Gebietsstandsaktualisierung Fälle auftreten, bei denen der Arbeitsagenturbezirk einer Person wechselt, ohne

	dass diese Person umgezogen ist. Einen Problemfall stellt hinsichtlich der Gebietsstandsaktualisierung allerdings Berlin dar: Die Berliner Arbeitsagenturbezirke wurden im Laufe der Jahre immer wieder neu zugeschnitten, was auch durch die Gebietsstandsumschlüsselung nicht vollständig bereinigt werden konnte. Zu beachten ist auch, dass bis circa 1989 in Berlin keine Zuordnung zu verschiedenen Arbeitsagenturen stattfand, sondern in den meisten Fällen AA 901 (= Landesarbeitsamt Berlin) für Berlin eingetragen war. **2) LHG, XLHG, ASU, XASU** Für diese Quellen ist das Merkmal nicht gebietsstandsbereinigt.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird lediglich das Gebiet der Regionaldirektion, in dem der Wohnort des Versicherten liegt, ausgewiesen.

5.6.3 Wohnort Regionaldirektion (wo_rd)

Variablenlabel	Wohnort Regionaldirektion
Variablenname	wo_rd
Kategorie	Ortsangaben
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Regionaldirektion | Arbeitsagentur
Detailbeschreibung	Das Merkmal beschreibt die Regionaldirektion des Wohnorts des BA-Kunden. Für die Quellen LHG und XLHG gilt der Wohnort für den Zeitraum des Satzes. Ein Wechsel des Wohnorts (bereits auf Ebene der Geschäftsstellen bzw. Gemeinden) löst einen Split des Satzes aus. Bei ASU und XASU enthält das Merkmal den Wohnort zum Beginn der Arbeitslosigkeit bzw. des Arbeitsuchendenstatus. Das Merkmal ist für die BeH und LeH ab 1999 vorhanden, für LHG, XLHG und XASU ab 2005. Für ASU Sätze ist das Merkmal bereits vor 1999 gefüllt. **1) BeH** Für die Quellen BeH und LeH ist das Merkmal gebietsstandbereinigt. **2) LHG, XLHG, ASU, XASU** Für diese Quellen ist das Merkmal nicht gebietsstandsbereinigt.

5.6.4 Wohnort Bundesland (wo_bula)

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

5.6.5 Arbeitsort Kreis (ao_kreis)

Variablenlabel	arbeitsort kreis
Variablenname	ao_kreis
Kategorie	Ortsangaben
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Bundesland | Kreis
Detailbeschreibung	Angegeben ist jeweils der Kreis (kreisfreie Stadt oder Landkreis), in dem der Betrieb des Beschäftigten angesiedelt ist. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des 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. Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die Kreisangaben auf den Gebietsstand vom 31.12.2008 umgeschlüsselt, d.h. in allen Kalenderjahren richtet sich die Zuordnung eines Betriebsortes zu einem Kreis nach den Grenzen, welche die Kreise zum 31.12.2008 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.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt.

5.6.6 Arbeitsort Arbeitsagentur (ao_aa)

Variablenlabel	Arbeitsort Arbeitsagentur
Variablenname	ao_aa
Kategorie	Ortsangaben
Herkunft	BeH, ASU
Datentyp	Numerisch
Hierarchie	Regionaldirektion | Arbeitsagentur
Detailbeschreibung	Das Merkmal enthält die Arbeitsagentur des Arbeitsortes. Im Gegensatz zum Merkmal „Wohnort Arbeitsagentur“ ist dieses Merkmal für BeH Sätze ab 1975, für ASU Sätze ab dem Jahr 2000 gefüllt. Das Merkmal ist nur für die Quelle BeH gebietsstandsbereinigt zum 31.12.2008.
Besonderheiten	Dieses Merkmal wird wegen seiner besonderen datenschutzrechtlichen Sensibilität nur auf Antrag und nur in begründeten Fällen unvergröbert bereitgestellt. Ansonsten wird lediglich das Gebiet der Regionaldirektion, in dem der Wohnort des Versicherten liegt, ausge-

	wiesen. Das Merkmal ist nur für die Quelle BeH gebietsstandsbereinigt zum 31.12.2008.

5.6.7 Arbeitsort Regionaldirektion (ao_rd)

Variablenlabel	Arbeitsort Regionaldirektion
Variablenname	ao_rd
Kategorie	Ortsangaben
Herkunft	BeH, ASU
Datentyp	Numerisch
Hierarchie	Regionaldirektion | Arbeitsagentur
Detailbeschreibung	Das Merkmal beschreibt die Regionaldirektion des Arbeitsortes des BA-Kunden. Im Gegensatz zum Merkmal „Wohnort Regionaldirektion“ ist dieses Merkmal für BeH Sätze ab 1975, für ASU Sätze ab dem Jahr 2000 gefüllt. Das Merkmal ist nur für die Quelle BeH gebietsstandsbereinigt zum 31.12.2008.

5.6.8 Arbeitsort Bundesland (ao_bula)

Variablenlabel	arbeitsort bula
Variablenname	ao_bula
Kategorie	Ortsangaben
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Bundesland | Kreis
Detailbeschreibung	Dieses Merkmal stellt eine Aggregation der Kreise zu den 16 deutschen Bundesländern dar.

5.7 Generierte technische Merkmale

5.7.1 Quelle des Satzes (quelle)

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

5.7.2 Satzzähler Konto (spell)

Variablenlabel	Satzzähler Konto
Variablenname	spell
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, LHG, XLHG, ASU, XASU

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 „Spellzähler Konto“ kann die ursprüngliche Sortierung auf einfache Art und Weise wieder hergestellt werden. Die Sortierung erfolgt grundsätzlich zunächst nach dem Beginndatum der Episode und dann nach der Quelle.

5.7.3 Satzzähler pro Episode (level2)

Variablenlabel	Satzzähler pro Episode
Variablenname	level2
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH , LHG, XLHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Zähler, der zeitlich parallel liegende Sätze zählt. Das Merkmal wird beim Episodensplitting erzeugt und bezieht sich auf die gesplitteten Sätze. Dieser Levelzähler zeigt an, ob sich innerhalb eines Kontos Sätze zeitlich überschneiden. Dies ist beispielsweise bei mehreren gleichzeitig bestehenden Beschäftigungsverhältnissen oder beim Zusammentreffen von Leistungsbezug und geringfügiger Beschäftigung der Fall. level2 nummeriert die Sätze einer solchen Episode, beginnend bei 0. Innerhalb der Episode ist zunächst nach Quellen sortiert (BeH, LeH, LHG, ASU, XLHG, XASU). Innerhalb der Quelle BeH stehen nicht-geringfügige Beobachtungen vor geringfügigen, danach ist absteigend nach Tagesentgelt sortiert. Innerhalb der Quelle LeH ist nach Leistungsart sortiert, innerhalb der anderen Quellen ist die Sortierung zufällig.

5.7.4 Satzzähler pro Episode und Quelle (level1)

Variablenlabel	Satzzähler pro Episode und Quelle
Variablenname	level1
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH , LHG, XLHG, ASU, XASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Mit Hilfe dieser Level-Variablen werden (zeitgleiche) Sätze pro Episode und Quelle gezählt. Das Merkmal wird beim Episodensplitting erzeugt und bezieht sich auf die gesplitteten Sätze. Dieser Levelzähler zeigt an, ob sich innerhalb eines Kontos Sätze zeitlich überschneiden. Level1 differenziert anders als Level2, der zeitlich paralle-

	le Sätze zählt zusätzlich auch noch nach der Quelle dieser Sätze. Die Level1-Variable somit nimmt jeweils für den ersten Datensatz, der je Episode und Quelle vorliegt den Wert 0 an. Weitere parallel verlaufende Datensätze für dieselbe Episode und Quelle werden dann hochgezählt. Innerhalb der Quelle BeH stehen nicht-geringfügige Beobachtungen vor geringfügigen, danach ist absteigend nach Tagesentgelt sortiert. Innerhalb der Quelle LeH ist nach Leistungsart sortiert, innerhalb der anderen Quellen ist die Sortierung zufällig. (Für jede Quelle gilt: level2 – level1 ≡ const; für Quelle BeH ist const = 0.)

5.7.5 Jahr (jahr)

Variablenlabel	Jahr (30. Juni)
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.
	
	`use siab_7508_v1.dta`
	`gen jahr = year(begepi)`
	`sort betnr jahr`
	`merge betnr jahr using siab_7508_v1_btr_basis.dta, uniqusing sort`

6 Literatur

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

Bundesagentur für Arbeit (Hg.) 1988: Klassifizierung der Berufe – Systematisches und Alphabetisches Verzeichnis der Berufsbenennungen, Nürnberg.

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

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

Drews, Nils 2006: Qualitätsverbesserung der Bildungsvariable in der IAB-Beschäftigtenstichprobe 1975-2001. (FDZ Methodenreport, 05/2006 (de)), Nürnberg, 16 S.

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

Meinken, Holger; Koch, Iris 2004: BA-Beschäftigtenpanel 1998-2002. Codebuch, Nürnberg.

Oberschachtsiek, Dirk; Scioch, Patrycja; Seysen, Christian; Heining, Jörg 2009: Stichprobe der Integrierten Erwerbsbiografien IEBS - Handbuch für die IEBS in der Fassung 2008. (FDZ Datenreport, 03/2009 (de)), Nürnberg.

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

7 Anhang

7.1 Bereinigungsschritte in der IABS

In Abschnitt 1.3 wurde bereits auf Unterschiede zwischen SIAB, IABS und IEBS eingegangen. An dieser Stelle wir dargestellt, welche Bereinigungsschritte der IABS bei der SIAB nicht mehr durchgeführt werden. Die Zahl der betroffen Datensätze in der IABS 2004 gibt einen Hinweis auf die Bedeutung der einzelnen Schritte.

Bereinigungen in der IABS	**Zahl der betroffenen Sätze in der IABS 2004**
Sätze gelöscht, falls Enddatum kleiner 1975	0 Sätze
Sätze gelöscht, falls Beginndatum größer 2004	7.923 Sätze
Sätze gelöscht, falls Abgabegrund == 35 (Arbeitskampf)	18 Sätze
Sätze gelöscht, falls Meldungen aus Ost-Deutschland vor 1992	182.347 Sätze
Sätze gelöscht, falls Abgabegrund 33 und Entgelt == 0	46.915 Sätze
Sätze gelöscht, falls Abgabegrund fehlt (nur LeH)	1.260 Sätze
Umsetzten des Beginndatums auf den 01.01.1975, falls Beginndatum kleiner 1975 ist	63 Sätze
Umsetzten des Enddatums auf den 31.12.2004, falls Enddatum größer 2004 ist	169 Sätze
Umsetzten des Tagesentgelts auf null, falls das Tagesentgelt kleiner 0,5 ist	

Korrektur bei mehrfachem Leistungsbezug	**Zahl der betroffenen Sätze in der IABS 2004**
Wenn eine Meldung vollständig innerhalb einer anderen, längeren Meldung liegt, wird die längere Meldung behalten und die kürzere gelöscht.	20.934 Sätze
Wenn zwei Meldungen gleich lang sind, wird die Meldung mit der höchsten Priorität behalten:	10 Sätze
Bei gleicher Priorität wird die Meldung mit dem höheren Leistungssatz behalten.	3.278 Sätze
Bei Teilüberschneidung wird die zeitlich frühere Meldung gekürzt.	6.980 Sätze

Korrektur der Überschneidungen von einem Tag	Zahl der betroffenen Sätze in der IABS 2004
Es werden alle Fälle korrigiert, die sich genau um einen Tag überschneiden. Bei der Korrektur wird folgendermaßen vorgegangen: - Erste Meldung hat Entgelt größer Null: Erste Meldung wird komplett behalten, alle weiteren Meldungen der Episode werden um einen Tag verkürzt. - Erste Meldung hat Entgelt gleich Null: Sind in der Episode nur BeH-Meldungen mit Entgelt gleich Null, dann wird der erste Satz behalten. Darauf folgende Meldungen werden um einen Tag verkürzt. Sind in der Episode noch Meldungen zu geringfügiger Beschäftigung vorhanden oder LeH-Meldungen, dann wird die erste dieser Meldungen komplett behalten. Alle anderen werden korrigiert. Die Korrektur wird nicht durchgeführt bei Meldungen mit Originaldauer von einem Tag.	69.259 Sätze davon Sozialversicherungspflichtig Beschäftigte: 19.946 Sätze Geringfügig Beschäftigte: 874 Sätze Leistungsbezug: 48.439 Sätze

Bereinigung bei Überschneidungen von BeH- und LeH - Meldungen	Zahl der betroffenen Sätze in der IABS 2004
Hierbei gilt, dass in der Regel die BeH-Meldung die LeH- Meldung schlägt. Ausnahmen sind BeH-Meldungen mit einem Entgelt gleich Null und BeH-Meldungen, die komplett im Zeitraum einer LeH-Meldung liegen. In diesen Fällen wird nur die BeH-Meldung gelöscht bzw. korrigiert. Episoden, in denen nur Überschneidungen von BeH-Meldungen zu geringfügiger Beschäftigung und LeH-Meldungen vorliegen, werden nicht korrigiert. Wird jedoch ein Satz gelöscht, werden alle weiteren Fälle der gleichen Episode und der gleichen Quelle, also auch geringfügige Beschäftigungsmeldungen gelöscht	Gelöschte Sätze: 536.754 Sätze Davon waren über 80 % LeH-Sätze.

Streikbereinigung	Zahl der betroffenen Personen in der IABS 2004
Wegen Aussperrung im Streik von Betrieben der Automobilindustrie in Baden-Württemberg und Hessen im Juni/Juli 1984 ergeben sich Beschäftigungsunterbrechungen. Daher wird für diese Personen eine Streikbereinigung durchgeführt. Es werden alle Sätze gelöscht, die in der Lücke liegen.	1.830 Personen

Ergänzungsverfahren	Zahl der betroffenen Sätze in der IABS 2004
In der IABS werden Sätze ergänzt, wenn folgende Bedingungen erfüllt sind: • Lücke zwischen zwei Datensätzen ist größer als sieben Tage • Satz vor der Lücke, bzw. letzter Datensatz im Konto einer Person beschreibt: o Sozialversicherungspflichtige Beschäftigung (nicht geringfügig), weist einen o Gültigen Eintrag im Merkmal blh_ao_kreis auf und o der Abgabegrund weist auf Beschäftigung hin (keine Abmeldung) • Satz nach der Lücke ist eine BeH – Meldung. • Jahresmeldungen am Ende eines Kontos werden nicht ergänzt. Die Länge des ergänzten Zeitraumes hängt davon ab, ob der Zeitraum ins alte oder ins neue Meldeverfahren fällt. Ist die Lücke kleiner als der maximal zu ergänzende Zeitraum, wird der Zeitraum komplett ergänzt. **Für alle ergänzten Meldungen werden das Entgelt, das Tagesentgelt und der Abgabegrund auf 0 gesetzt**, das Episodenanfangs- und Enddatum werden entsprechend der zu ergänzenden Zeiträume gesetzt. Die restlichen Merkmale werden von der Meldung vor der Lücke übernommen.	Anzahl Konten insgesamt: 1.293.819 Sätze Anzahl der Konten mit mind. einem Ergänzungsfall: 397.214 Sätze Ergänzt insgesamt (Summe btyp = 2,4,5,6): 725.967 Sätze

7.2 Auszählungen

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

8 Abkürzungsverzeichnis

AA	Agentur für Arbeit / Arbeitsamt
ALG	Arbeitslosengeld
ARGE	Arbeitsgemeinschaft
ASU	Arbeitsuchende-Historik
BA	Bundesagentur für Arbeit
BeH	Beschäftigten-Historik
BfA	Bundesversicherungsanstalt für Angestellte
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 –
FDZ	Forschungsdatenzentrum der Bundesagentur für Arbeit am IAB
IAB	Institut für Arbeitsmarkt- und Berufsforschung
IABS	IAB-Beschäftigtenstichprobe
IEB	Integrierte Erwerbsbiografien
IEBS	Stichprobe der Integrierten Erwerbsbiografien
LeH	Leistungsempfänger-Historik
LHG	Leistungs-Historik Grundsicherung
LVA	Landesversicherungsanstalt
MTH	Maßnahmeteilnehmer-Historik
SchwbG	Gesetz zur Sicherung der Eingliederung Schwerbehinderter in Arbeit, Beruf und Gesellschaft – Schwerbehindertengesetz-
SGB	Sozialgesetzbuch
SIAB	Stichprobe der Integrierten Arbeitsmarktbiografien
VerBIS	Vermittlungs- und Beratungsinformationssystems
XASU	Arbeitsuchenden-Historik aus XSozial-BA-SGB II
XLHG	Leistungsempfänger-Historik Grundsicherung aus XSozial-BA-SGB II

Impressum

FDZ-Datenreport 1/2010

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

Redaktion
Stefan Bender, Britta Hübner

Technische Herstellung
Britta Hübner

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

Bezugsmöglichkeit
http://doku.iab.de/fdz/reporte/2010/DR_01-10.pdf

Internet
http://fdz.iab.de/

Rückfragen zum Inhalt an:
Jörg Heining
Forschungsdatenzentrum
Regensburger Str. 104
90478 Nürnberg
Telefon: 0911 / 179-5392
E-Mail: Joerg.Heining@iab.de