

Forschungsdatenzentrum

der Bundesagentur für Arbeit  
im Institut für Arbeitsmarkt-  
und Berufsforschung

FDZ

# FDZ-Datenreport

Dokumentation zu Arbeitsmarktdaten

05/2013

DE

## Stichprobe der Integrierten Arbeitsmarktbiografien

Campus File 7508 (SIAB-CF 7508)

Version 2

Aktualisierte Version vom 08.08.2013

Matthias Umkehrer,  
Daniel Werner



Bundesagentur für Arbeit

# Stichprobe der Integrierten Arbeitsmarktbiografien

Campus File 7508 (SIAB-CF 7508)

Matthias Umkehrer (IAB)

Daniel Werner (IAB)

Dokumentation: Version 2 (SIAB-CF\_7508\_v2\_dok1)

Datensatz: Version 2 (SIAB-CF\_7508\_v2)

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 has a dual function: on the one hand, users can ascertain whether the data are suitable for their research task; on the other, the reports can be used to prepare the analyses.

# Inhaltsverzeichnis

Zusammenfassung . . . . .	5
Abstract . . . . .	5
1 Einleitung . . . . .	6
2 Veränderung zu früheren Versionen . . . . .	7
3 Datenquellen . . . . .	7
3.1 Beschäftigten-Historik (BeH) . . . . .	7
3.2 Leistungsempfänger-Historik (LeH) . . . . .	8
3.3 Leistungs-Historiken Grundsicherung (LHG/XLHG) . . . . .	8
3.4 Arbeitsuchenden-Historiken (ASU/XASU) . . . . .	10
4 Generierung des SIAB-CF 7508 . . . . .	10
4.1 Datenschutzrechtlicher Hintergrund . . . . .	10
4.2 Datenselektion . . . . .	10
4.3 Datensatzkonzeption . . . . .	11
4.4 Stichprobenziehung . . . . .	12
4.5 Anonymisierung . . . . .	13
5 Merkmalspektrum . . . . .	15
5.1 Identifikatoren . . . . .	15
5.1.1 Systemfreie Personennummer (persnr) . . . . .	15
5.2 Gültigkeitszeitraum . . . . .	16
5.2.1 Beginndatum Originalsatz (begorig) . . . . .	16
5.2.2 Enddatum Originalsatz (endorig) . . . . .	16
5.2.3 Beginndatum der Episode (begepi) . . . . .	16
5.2.4 Enddatum der Episode (endepe) . . . . .	17
5.3 Informationen zur Person . . . . .	17
5.3.1 Geschlecht (frau) . . . . .	17
5.3.2 Geburtsjahr (gebjahr) . . . . .	18
5.3.3 Ausbildung (bild) . . . . .	18
5.3.4 Schulausbildung (schbild) . . . . .	20
5.4 Informationen zu Beschäftigung, Leistungsbezug und Arbeitssuche . . . . .	20
5.4.1 Abgabegrund/ Beendigungsgrund/ SGB II-Einstellungsgrund/ Abmeldegrund (grund_gr) . . . . .	20
5.4.2 Tagesentgelt, täglicher Leistungssatz (tentgelt_gr) . . . . .	22
5.4.3 Stellung im Beruf und Arbeitszeit (stib) . . . . .	24
5.4.4 Erwerbsstatus (erwstat_gr) . . . . .	25
5.5 Generierte technische Merkmale . . . . .	26
5.5.1 Quelle des Satzes (quelle_gr) . . . . .	26
5.5.2 Satzzähler pro Episode (level2) . . . . .	27
5.5.3 Satzzähler pro Episode und Quelle (level1) . . . . .	27
5.6 Neu gebildete Merkmale im SIAB-CF 7508 . . . . .	28
5.6.1 Tage in soz.pfl. Besch. in den 10 Jahren vor Arbeitslosengeldbezug (tage_svb_gr) . . . . .	28

5.6.2	Tage in geringf. Besch. in den 10 Jahren vor Arbeitslosengeldbezug (tage_grfb_gr) . . . . .	29
5.6.3	Tage in Leistungsbezug in den 10 Jahren vor Arbeitslosengeldbezug (tage_leist_gr) . . . . .	29
5.6.4	Arbeitsort Ost/West (ao_ost) . . . . .	30
5.6.5	Arbeitsort letzte Besch. vor Arbeitslosengeldbezug Ost/West (letzte_besch_ost) . . . . .	30
5.7	Zusätzliche Merkmale im SIAB-R 7508 . . . . .	31
5.7.1	Betriebsnummernzähler (bnn) . . . . .	31
5.7.2	Staatsangehörigkeit (deutsch) . . . . .	31
5.7.3	Beruf – ausgeübte Tätigkeit (beruf_gr) . . . . .	32
5.7.4	Beschäftigungsdauer vor dem Alter von 17 (tage_jung) . . . . .	33
5.7.5	Beschäftigungsdauer nach dem Alter von 62 (tage_alt) . . . . .	33
5.7.6	Wirtschaftszweig 73 – Wirtschaftsgruppe 73 (w73_gen_gr) . . . . .	33
5.7.7	Wirtschaftszweig 03 – Wirtschaftsgruppe 03 (w03_gr) . . . . .	34
5.7.8	Arbeitsort Region (ao_region) . . . . .	35
5.7.9	Pendlerstatus (pendler) . . . . .	35
5.7.10	Satzzähler Konto (spell) . . . . .	36
6	Arbeiten mit dem Campus File . . . . .	37
7	Abschließende Hinweise und Datenzugang . . . . .	37
	Literatur . . . . .	39

## Zusammenfassung

Ausschließlich für die Nutzung in der akademischen Lehre an Universitäten, Hochschulen oder Forschungsinstituten hat das Institut für Arbeitsmarkt- und Berufsforschung (IAB) einen Campus File (CF) als absolut anonymisierte Stichprobe aus dem Scientific Use File der Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB-R 7508) erzeugt, das SIAB-CF 7508. Dieser Datenreport gibt einen kurzen Überblick über dessen Konzeption und Erstellung sowie Möglichkeiten und Einschränkungen im Umgang mit dem Campus File.

## Abstract

Exclusively for use in academic teaching at universities or research institutes, the Institute for Employment Research (IAB) generated absolutely anonymized data, a so-called campus file (CF), based on the data of the Scientific Use File of the Sample of Integrated Labour Market Biographies (SIAB Regional File 7508), the SIAB-CF 7508. This documentation provides a brief description of the data and its limitations.

**Stichwörter: Deutsche administrative Mikrodaten, Arbeitsmarktdaten, Datensatzbeschreibung, Absolut anonymisierte Daten, Campus File**

**Danksagung: Für wertvolle Anregungen und Kommentare während der Erstellung des Datensatzes sowie dessen Dokumentation danken wir Annette Bergemann und den Kolleginnen und Kollegen des Forschungsdatenzentrums der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (FDZ), insbesondere Stefan Bender, Dana Müller und Alexandra Schmucker. Für den vorliegenden Datenreport wurden Teile des Datenreports des Regionalfile der Stichprobe der Integrierten Arbeitsmarktbiografien 1975-2008 (SIAB-R 7508) (FDZ-Datenreport 07/2011) sowie des PASS Campus Files des Panels "Arbeitsmarkt und soziale Sicherung" (FDZ-Datenreport 06/2011) übernommen. Die Verantwortung für den Inhalt und mögliche Fehler verbleibt einzig bei den Autoren des vorliegenden Datenreports.**

# 1 Einleitung

Das Campus File (CF) der Stichprobe der Integrierten Arbeitsmarktbiographien (SIAB) (SIAB-CF 7508) wird als (disproportional gezogene) Stichprobe aus dem Scientific Use File der Stichprobe der Integrierten Arbeitsmarktbiografien 1975-2008 (SIAB-R 7508) erstellt. Dieses wiederum stellt eine faktisch anonymisierte Version aus der nur schwach anonymisierten Stichprobe der Integrierten Arbeitsmarktbiografien 1975-2008 (SIAB 7508) dar. Letztere entspricht einer zwei Prozent Stichprobe aus den Integrierten Erwerbsbiografien (IEB) des Instituts für Arbeitsmarkt- und Berufsforschung (IAB), welche aus der Grundgesamtheit aller Personen erstellt wird, die zwischen 1975 und 2008 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 als arbeitsuchend gemeldet (erfasst ab 2000),
- (Geplante) Teilnahme an arbeitsmarktpolitischer Maßnahme (erfasst ab 2000).

Diese Informationen stammen aus unterschiedlichen administrativen Datenquellen und werden tagesgenau abgebildet. Sie werden in den IEB für nicht-kommerzielle wissenschaftliche Zwecke aufbereitet und zusammengeführt.<sup>1</sup> Um das wissenschaftliche Arbeiten mit dieser Art von Daten im Rahmen der Lehre vermitteln zu können, soll das SIAB-CF 7508 die grundlegenden Strukturen der zugrunde liegenden Datensätze beibehalten aber gleichzeitig der Vorgabe absoluter Anonymität entsprechen. Das SIAB-CF 7508 darf ausschließlich in der wissenschaftlichen Lehre an Universitäten, Hochschulen oder Forschungsinstituten eingesetzt und in diesem Rahmen frei verwendet werden. Es lassen sich keinerlei valide inhaltliche Aussagen auf Basis des SIAB-CF 7508 ableiten.

Das SIAB-CF 7508 liefert Informationen zu den Erwerbsbiographien von 6.741 Personen, die im Jahr 2005 in den Leistungsbezug nach dem Rechtskreis SGB III (Arbeitslosengeld I) eintreten und zu diesem Zeitpunkt zwischen 25 und 50 Jahre alt sind. Für diese Personen beinhaltet das SIAB-CF 7508 detaillierte Erwerbshistorien beginnend mit der ersten Meldung eines Arbeitslosengeldbezugs (Arbeitslosengeld I) im Jahr 2005 bis zum Ende des Beobachtungszeitraums 2008. Informationen zum Erwerbsverlauf vor dem ersten Bezug von Arbeitslosengeld I im Jahr 2005 werden im SIAB-CF 7508 in vergrößerter Form abgebildet.

Der vorliegende Datenreport erläutert zuerst die Unterschiede zu früheren Versionen des SIAB-CF 7508 (Abschnitt 2). Abschnitt 3 beschreibt die dem SIAB-CF 7508 zugrunde liegende Datenbasis. Anschließend gibt der Datenreport einen Überblick zur Erstellung und

<sup>1</sup> Siehe Dorner/König/Seth (2011) für eine Dokumentation des SIAB-R 7508. Der vorliegende Datenreport baut überwiegend auf dieser Literatur auf und setzt diese auch voraus.

Konzeption des SIAB-CF 7508 (Abschnitt 4). Abschnitt 5 beschreibt die im SIAB-CF 7508 enthaltenen Merkmale. Abschnitt 6 diskutiert Möglichkeiten und Einschränkungen im Arbeiten mit dem Campus File und Abschnitt 7 erläutert den Datenzugang. Um sinnvoll mit dem Campus File arbeiten zu können, sollten stets die detaillierteren Datenbeschreibungen des SIAB-R 7508 (siehe Dorner/König/Seth 2011) herangezogen werden.

## 2 Veränderung zu früheren Versionen

Wie sich gezeigt hat, kommt es zum 1. Januar 2005 zu einer ungewöhnlichen Häufung von Zugängen in den Bezug von Arbeitslosengeld I. Weiterhin gehen diese Zugänge häufig mit einer Abmeldung aus dem Bezug von Arbeitslosengeld I zum 31. Dezember 2004 ohne bzw. mit ungültigem Abmeldegrund einher. In diesen Fällen kann nicht davon ausgegangen werden, dass zum 1. Januar 2005 ein neuer Eintritt in den Bezug von Arbeitslosengeld I erfolgt. Es scheint sich hier um eine Fortsetzung der Periode des Bezugs von Arbeitslosengeld I aus dem Vorjahr zu handeln.

In der ersten Version des SIAB-CF 7508 wurde der Beginn einer neuen Periode des Bezugs von Arbeitslosengeld I mit einem Eintritt in den Bezug von Arbeitslosengeld I gleichgesetzt. Die zweite Version des SIAB-CF 7508 unterscheidet sich von der Vorgängerversion darin, dass bei der Identifikation der Eintritte Unterbrechungen von weniger als sieben Tagen zwischen zwei Perioden des Bezugs von Arbeitslosengeld I als unschädlich angesehen werden. Liegen weniger als sieben Tage zwischen zwei Perioden, so werden sie als eine einzige Periode angesehen. Nach dieser neuen Abgrenzung bei der Identifikation der Eintrittskohorte fällt die Anzahl der Personen, für die im Jahr 2005 eine Periode des Bezugs von Arbeitslosengeld I beginnt, geringer aus und damit auch die Zahl der Personen im SIAB-CF 7508. Auch reduziert sich das Merkmalspektrum gegenüber der Vorgängerversion (siehe Abschnitt 4.5).

## 3 Datenquellen

Die Informationen in den IEB beruhen auf Daten, die sich zum einen aus dem Meldeverfahren zur Sozialversicherung generieren und in der Beschäftigten-Historik (BeH) zusammengeführt werden. Zum anderen entstammen sie den Geschäftsprozessen der Bundesagentur für Arbeit (BA) sowie der SGB II-Träger. Letztere Informationen münden in der Leistungsempfänger-Historik (LeH), den Leistungs-Historiken Grundsicherung (LHG/XLHG), den Arbeitssuchenden-Historiken (ASU/XASU) sowie der Maßnahmenteilnehmer-Historik (MTH). Mit Ausnahme der MTH werden Datensätze aus all diesen Quellen in die SIAB übernommen. Diese Datenquellen werden im Folgenden näher beschrieben.

### 3.1 Beschäftigten-Historik (BeH)

Informationen zu sozialversicherungspflichtiger Beschäftigung entstammen der Beschäftigten-Historik des IAB. Diese wiederum beruht auf Daten aus dem Meldeverfahren zur

Sozialversicherung. Dessen rechtliche Grundlage ist das mit Wirkung zum 1. Januar 1973 eingeführte (und seit dem 1. Januar 1991 auf Ostdeutschland ausgedehnte) integrierte Meldeverfahren zur Kranken-, Renten- und Arbeitslosenversicherung, gesetzlich festgeschrieben in der Datenerfassungs- und -übermittlungsverordnung (DEÜV), vormals Datenerfassungsverordnung (DEVO) und Datenübermittlungsverordnung (DÜVO) (für weitere Details siehe Wermter/Cramer 1988; Bender u. a. 1996, S. 4 ff.; Schermer 1999). Es verlangt von den Arbeitgebern Meldungen für alle sozialversicherungspflichtig beschäftigten Arbeitnehmer an die zuständigen Sozialversicherungsträger. Die BeH umfasst alle Arbeiter und Angestellten sowie alle Auszubildenden, soweit sie nicht von der Sozialversicherungspflicht befreit sind. Dies bedeutet, dass Informationen über Beamte, Selbstständige und ordentliche Studierende grundsätzlich nicht in der BeH enthalten sind (vgl. Wermter/Cramer 1988). Seit der Änderung des Meldeverfahrens zum 1. Januar 1999 werden auch geringfügig Beschäftigte und mithelfende Familienangehörige erfasst (in den Daten erst ab 1. April 1999 enthalten).

Jedes Jahr, in dem ein Beschäftigungsverhältnis für eine Person besteht, ist durch mindestens eine Meldung abgebildet. Die Daten werden von den Krankenkassen aufgenommen, an die Deutsche Rentenversicherung übermittelt und dann an die Bundesagentur für Arbeit weitergeleitet. Dort werden sie in einer laufenden Datei gesammelt und anschließend vom IAB in seine Historik-Datei integriert.

### **3.2 Leistungsempfänger-Historik (LeH)**

Die Leistungsempfänger-Historik des IAB erfasst Zeiträume, in denen Personen Lohnersatzleistungen beziehen. Die Leistungen beinhalten Arbeitslosengeld, Arbeitslosenhilfe und Unterhaltsgeld, also ausschließlich Leistungen nach dem Rechtskreis SGB III. Da der Leistungsanspruch von der Erfüllung gesetzlicher Voraussetzungen abhängt, werden Arbeitslosigkeitsperioden, in denen die Voraussetzungen nicht vorliegen (z.B. fehlende Bedürftigkeit im Falle der Arbeitslosenhilfe oder Nichterfüllung der Anwartschaftszeit beim Arbeitslosengeld) nicht in der LeH berichtet.

### **3.3 Leistungs-Historiken Grundsicherung (LHG/XLHG)**

Während die LeH die Leistungen nach dem Rechtskreis SGB III abbildet, beinhalten LHG und XLHG den Leistungsbezug gemäß SGB II. Dieser umfasst sowohl Leistungen zur Sicherung des Lebensunterhalts (z.B. Arbeitslosengeld II) wie auch Zuschläge zum Arbeitslosengeld oder zusätzliche Leistungen. Im Gegensatz zu den Leistungen aus dem Rechtskreis des SGB III ist die Bundesagentur für Arbeit nicht mehr zwingend alleiniger Träger dieser Leistungen. Es wird demnach grundsätzlich nach drei möglichen Trägerarten unterschieden:

- Arbeitsgemeinschaften (ARGE) aus gemeinsamer Aufgabenwahrnehmung von BA und Kommune.



- Bei getrennter Trägerschaft bleibt die Aufgabenteilung zwischen BA und Kommune bestehen.<sup>2</sup>
- 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-Historiken Grundsicherung umfassen 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-SGBII“ (XLHG) Daten erfasst, welche nach dem Übermittlungsstandard X-Sozial gemeldet wurden. In der Regel findet A2LL bei ARGen Verwendung, während der Standard X-Sozial von den optierenden Kommunen genutzt wird. Bei Kommunen mit getrennter Trägerschaft kommen beide Verfahren zum Einsatz.

Die Ermittlung der Höhe des Leistungsbezugs erfolgt nicht auf individueller Ebene, 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 der LHG und der XLHG sind grundsätzlich alle gemeldeten Berechtigten nach § 7 SGB II, also erwerbsfähige Hilfebedürftige und ggf. ihre Bedarfsgemeinschaftsmitglieder enthalten.

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

- an transferrelevanten Geburtstagen (14, 15, 18, 25 und 65) 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.

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

### 3.4 Arbeitssuchenden-Historiken (ASU/XASU)

Daten über Arbeitssuchende werden in der Arbeitssuchenden-Historik (ASU) gespeichert. Die Quelle ASU enthält dabei Informationen zu Arbeitssuchenden, 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 arbeitssuchenden Personen mit Bezug von Arbeitslosengeld II von zugelassenen kommunalen Trägern enthalten, welche nach dem X-Sozial-Standard gemeldet werden.

## 4 Generierung des SIAB-CF 7508

### 4.1 Datenschutzrechtlicher Hintergrund

Die dem SIAB-CF 7508 zugrunde liegenden Daten entstammen zum einen dem Meldeverfahren zur Sozialversicherung und zum anderen den Geschäftsprozessen der Bundesagentur für Arbeit. Da es sich hierbei um Pflichtangaben handelt, um z.B. die Beiträge und späteren Ansprüche für die Sozialversicherungen zu berechnen, gelten sie als Sozialdaten (§ 67 SGB X) und unterliegen somit einem besonders strikten Datenschutz, dem Sozialgeheimnis (§ 35 SGB I). Der Gesetzgeber hat jedoch Möglichkeiten zur Nutzung dieser Daten im Rahmen der wissenschaftlichen Forschung im Sozialleistungsbereich geschaffen (§ 282 Abs. 7 SGB III, § 75 SGB X). Auf dieser rechtlichen Basis kann das FDZ faktisch anonymisierte Scientific Use Files an externe Forscherinnen und Forscher übermitteln und weniger stark anonymisierte Mikrodaten im Rahmen von Gastaufenthalten bzw. der Datenfernverarbeitung zugänglich machen.<sup>3</sup> Diesen Datenzugangswegen sind jedoch strenge Grenzen gesetzt, sodass eine Weitergabe der Daten für die breite Öffentlichkeit oder Lehrveranstaltungen nicht möglich ist (vgl. Hochfellner u. a. 2012).

Um dennoch Nachwuchswissenschaftlerinnen und -wissenschaftler frühzeitig mit den komplexen Daten des FDZ vertraut zu machen, wurde ein Campus File erstellt, das absolute Anonymität gewährleistet. Damit können nun mikroökonomische Methoden angewendet, jedoch keinerlei valide inhaltliche Aussagen abgeleitet werden.

### 4.2 Datenselektion

Das SIAB-CF 7508 basiert auf dem SIAB-R 7508. Dieses bildet die Erwerbsverläufe von 1.515.463 Individuen ab, deren Erwerbsbiografien in 34.862.777 Datenzeilen dokumentiert sind. Hieraus werden die Erwerbshistorien aller Personen selektiert, die im Jahr 2005 zwischen 25 und 50 Jahre alt sind. Diese Auswahl wird wiederum auf diejenigen Personen beschränkt, bei denen der Beginn einer Episode des Bezugs von Arbeitslosengeld I im

<sup>3</sup> Daten können als faktisch anonym angesehen werden wenn unverhältnismäßig viel Zeit, Kosten und Arbeit aufgewendet werden müssen, um einen Merkmalsträger im Datensatz zu deanonymisieren. Die Kosten einer Deanonymisierung müssen deren Nutzen also übersteigen.

Jahr 2005 verzeichnet wird. Nach den unten beschriebenen Anonymisierungsschritten verbleiben noch 33.707 Personen im Datensatz. Diese Personen bilden die Grundgesamtheit für die abschließende Stichprobenziehung.

### 4.3 Datensatzkonzeption

Die besondere Herausforderung bei der Erstellung eines Campus Files liegt darin, auf der einen Seite absolute Anonymität in den Daten sicherzustellen, was meist mit einer erheblichen Reduzierung des Informationsgehalts einhergeht. Auf der anderen Seite sollen die Daten inhaltlich sinnvolle und interessante Analysen ermöglichen. Mit der Beschränkung auf Personen, die 2005 eine Periode des Bezuges von Arbeitslosengeld I beginnen, wird beiden Anforderungen Rechnung getragen. Die so abgegrenzte Stichprobe ist leichter zu anonymisieren und bietet gleichzeitig eine Grundlage für das Anwenden von Methoden wie der Verweildaueranalyse. Des Weiteren soll anhand des SIAB-CF 7508 die Vermittlung von Techniken zur mikroökonomischen Wirkungsanalyse von Maßnahmen der aktiven Arbeitsmarktpolitik möglich sein. Wie zuvor erwähnt sind Informationen aus der Maßnahme-Teilnahme-Historik (MTH) in der SIAB allerdings grundsätzlich nicht enthalten. Es besteht aber die Möglichkeit Maßnahmeteilnehmer auf Basis des Abgangsgrundes aus Arbeitslosigkeit zu identifizieren. Hierbei ergeben sich für eine mikroökonomische Wirkungsanalyse folgende Einschränkungen:

- Aus datenschutzrechtlichen Gründen fehlen die Informationen aus der Arbeitsuchenden-Historik (ASU) für die Jahre 2005 und 2006. Eine Wirkungsanalyse von aktiver Arbeitsmarktpolitik auf Basis des Abgangsgrundes aus Arbeitslosigkeit ist somit nur für die Jahre 2007 und 2008 möglich.
- Auf Basis der Abgangsgründe aus Arbeitslosigkeit ist es zwar möglich den Beginn einer Maßnahme im SIAB-R 7508 bzw. im SIAB-CF 7508 zu identifizieren. Es gestaltet sich jedoch schwierig auf Basis dieser beiden Versionen der SIAB das Ende einer Maßnahme zu ermitteln. Dies gilt insbesondere für Maßnahmen, die in Verbindung mit einer (sozialversicherungspflichtigen) Beschäftigung stehen. In diesem Fall kann der Zeitpunkt des Übergangs von einer geförderten in eine ungeforderte Beschäftigung — und damit das Ende der Maßnahme — nicht immer eindeutig ermittelt werden. Im SIAB-CF 7508 erscheint daher nur eine Evaluation der Teilnehmer an Förderung der beruflichen Weiterbildung (FbW) und Deutschsprachlehrgängen (DSL) sinnvoll. Arbeitslose, die an solchen Maßnahmen teilnehmen, lassen sich anhand des Abgangsgrundes “FbW und Deutschsprachlehrgänge” (**grund\_gr** = 23) identifizieren.

Beim Arbeiten mit dem SIAB-CF 7508 im Allgemeinen und bei der Ermittlung des Beginns von Maßnahmen im Speziellen ist zu beachten, dass es sich beim SIAB-CF 7508 um einen “gesplitteten” Datensatz handelt. D.h.: “Falls sich innerhalb eines Kontos Sätze aus den Quellen überlappen, werden diese durch künstliche Sätze mit neuen Datumsangaben derart ersetzt, dass vollständig parallele bzw. überschneidungsfreie Zeiträume entstehen (Episodensplitting).” (Dorner/König/Seth 2011, S. 26). Dieses Vorgehen ist beispielhaft in Abbildung 1 dargestellt.

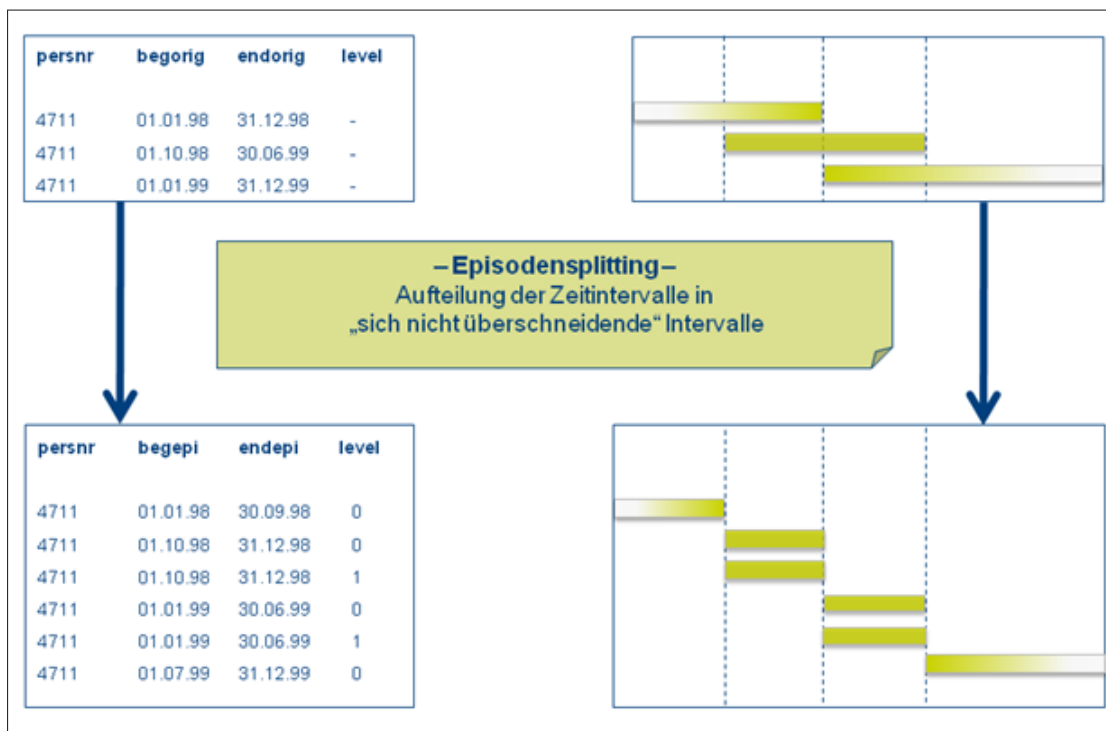


Abbildung 1: Episodensplitting

Anhand eines Vergleichs der ursprünglichen Datumsvariablen für Beginn und Ende des Originalsatzes (begorig und endorig) mit den neuen Datumsvariablen für Beginn und Ende eines Episodenzeitraums (begepi und endepe) lässt sich erkennen, ob ein Satz gesplittet wurde. Zur Wiederherstellung der Originaldaten können alle Sätze selektiert werden, bei denen der Beginn des Originalsatzes dem Beginn der gesplitteten Episode entspricht (begepi = begorig).<sup>4</sup>

#### 4.4 Stichprobenziehung

Das SIAB-CF 7508 entspricht einer disproportional gezogenen 20 Prozent Zufallsstichprobe aller gemäß Abschnitt 4.2 aus dem SIAB-R 7508 selektierten Personen. Unter diesen 33.707 Personen finden sich lediglich 1.264 Personen (3,7 Prozent), die an den Maßnahmen Förderung der beruflichen Weiterbildung (FbW) oder Deutschsprachlehrgängen (DSL) teilnehmen. Um mikroökonomische Wirkungsanalysen für Teilnehmer an FbW und DSL mittels des SIAB-CF 7508 zu ermöglichen, wird eine 50 Prozent Zufallsstichprobe aus den 1.264 Teilnehmern an FbW oder DSL gezogen. Diese umfasst 632 Personen. Die übrigen 6.109 Personen für das SIAB-CF 7508 werden als 18,8 Prozent Zufallsstichprobe aus den 32.443 verbleibenden Personen gezogen. Somit ist im SIAB-CF 7508 die Gruppe der Teilnehmer an FbW und DSL mit 9,4 Prozent deutlich stärker vertreten als in der Grundgesamtheit.

<sup>4</sup> Zum Episodensplitting des SIAB-R 7508 vgl. Abschnitt 3.5 in Dorner/König/Seth (2011).

## 4.5 Anonymisierung

Um die Anonymisierung von Datensätzen zu gewährleisten wird in der Regel auf zwei zentrale Strategien zurück gegriffen: die Informationsreduktion und die Datenveränderung (vgl. Südfeld 1987; Ronning u. a. 2005; Zwick 2008). Unter die Verfahren zur Datenveränderung fallen beispielsweise

- auf der Ebene der Merkmalsträger das "Klonen", also Vervielfältigen von Merkmalsträgern oder Zensierungs- (top/bottom-coding) sowie Ersetzungs-Verfahren (Replacement), wie z.B. Ausreißer mit dem Mittelwert zu ersetzen,
- auf der Ebene der Merkmale Vertauschungsverfahren (Swapping), Post-Randomisierung oder auch SAFE-Verfahren, Imputationsverfahren oder Stochastische Überlagerung (Überlagerung mit Zufallsfehlern) und
- auf der Ebene der Merkmalsausprägungen Mikroaggregationsverfahren.<sup>5</sup>

Das SIAB-R 7508, welches dem SIAB-CF 7508 zugrunde liegt, ist bereits faktisch anonymisiert. Dies wurde in erster Linie durch ausgewählte Löschungen sowie Aggregationen von datenschutzrechtlich sensiblen Informationen erreicht. Insbesondere wurden Angaben zur Quellenherkunft bestimmter Meldungen zusammengefasst sowie Angaben zum Beginn- bzw. Enddatum ausgewählter Meldungen angepasst. Darüber hinaus wurden Altersangaben zensiert und Merkmalsausprägungen von Variablen mit hohem Deanonymisierungsrisiko aggregiert oder gelöscht. Löschungen betreffen in erster Linie Informationen zu Leistungsempfang und Arbeitsuche während die beschäftigungsbezogenen Merkmale weitgehend erhalten geblieben und lediglich vergrößert worden sind.<sup>6</sup> Für das SIAB-CF 7508 wurde weiterhin geprüft, ob Ausprägungen bestimmter Variablen als kritisch oder sensibel betrachtet werden müssen. Dies wäre der Fall, wenn einige Ausprägungen einer Variablen nur selten im Datensatz zu beobachten sind. Auch wenn Beobachtungen mit bestimmten Ausprägungen vergleichsweise häufig im Datensatz auftreten, kann nicht ausgeschlossen werden, dass sich diese Beobachtungen auf nur wenige Personen konzentrieren. Eine Prüfung, bei der nur Beobachtungen alleine betrachtet werden, würde daher zu kurz greifen. Aus diesem Grund wurde geprüft, auf wie viele Personen sich die Merkmalsausprägungen einer Variable verteilen. Wenn bestimmte Merkmalsausprägungen nur für wenige Personen beobachtbar sind, werden die gesamten Erwerbshistorien dieser Personen vollständig aus dem Datensatz gelöscht. Merkmalsausprägungen werden hierbei als kritisch betrachtet, wenn weniger als 100 Personen diese aufweisen. Bei einer Stichprobenziehung von 20 Prozent könnten die daraus resultierenden Fallzahlen für diese Merkmalsausprägungen aus datenschutzrechtlicher Sicht zu gering sein.

Das SIAB-CF 7508 enthält die detaillierte Erwerbsbiografie einer Person beginnend mit dem Bezug von Arbeitslosengeld I im Jahr 2005 bis Ende des Beobachtungszeitraums 2008. Daher wurde die Prüfung für diesen Zeitraum durchgeführt. Weiterhin erfolgt die Prüfung sowohl vor als auch nach der Stichprobenziehung. Wenn sich eine bestimmte

<sup>5</sup> Eine detaillierte Beschreibung der einzelnen Verfahren findet sich bei Ronning u. a. (2005), S. 55 ff.).

<sup>6</sup> Für weitere Details und zum Konzept der faktischen Anonymität siehe Abschnitt 3.3 in Dörner/König/Seth (2011).

Merkmalsausprägung erst nach der Stichprobenziehung als kritisch erweist, dann wurden alle Personen aus der Grundgesamtheit, die dieses Merkmal aufweisen, gelöscht und eine erneute Stichprobenziehung durchgeführt. Somit werden Personen immer nur vor der Stichprobenziehung aus dem Datensatz gelöscht und nie danach.

Die beschriebene Prüfung wurde für folgende Variablen durchgeführt: Bildung (**bild**), Schulbildung (**schbild**), Abgabegrund (**grund\_gr**), Erwerbsstatus (**erwstat\_gr**) und Stellung im Beruf (**stib**) (vgl. die Beschreibung der Merkmale im SIAB-CF 7508 in Abschnitt 5). Bei der Variable **bild** wurden keine kritischen Merkmalsausprägungen identifiziert. Personen mit den folgenden Merkmalsausprägungen der übrigen Variablen wurden im Datensatz gelöscht und sind folglich nicht im SIAB-CF 7508 enthalten:

<b>Schulbildung (schbild)</b>	<ul style="list-style-type: none"> <li>- Abschluss der Schule für Lern- oder Geistigbehinderte [2]</li> <li>- Qualifizierter Hauptschulabschluss [4]</li> <li>- Ohne Fachhochschul-/Hochschulabschluss [12]</li> <li>- Fachhochschulabschluss/Diplom [13]</li> <li>- Hochschulabschluss/Diplom [14]</li> </ul>
<b>Abgabegrund (grund_gr)</b>	<ul style="list-style-type: none"> <li>- Wehr-/ Zivildienst/ Wehrübung/ Soldat [28]</li> <li>- Betriebliche Ausbildung [32]</li> <li>- Umzug [34]</li> <li>- Ausscheiden aus Erwerbsleben [36]</li> <li>- Wehr-/ Zivildienst/ Wehrübung [51]</li> <li>- Übergangsgeld [52]</li> <li>- Ausbildung/Studium [53]</li> <li>- Eigene Abmeldung [55]</li> <li>- Ende Gewährungszeitraum [56]</li> <li>- Altersrente/Grundsicherung im Alter [57]</li> <li>- Wegfall der Erwerbsfähigkeit [58]</li> </ul>
<b>Erwerbsstatus (erwstat_gr)</b>	<ul style="list-style-type: none"> <li>- Altersteilzeit [4]</li> <li>- Praktikanten [5]</li> <li>- Unständige Beschäftigung [6]</li> <li>- Sonstiger Beschäftigungsstatus [7]</li> <li>- Unterhaltsgeld [13]</li> <li>- Ratsuchend [25]</li> </ul>
<b>Stellung im Beruf (stib)</b>	<ul style="list-style-type: none"> <li>- Beschäftigung von zu Hause aus [7]</li> </ul>

Anmerkung: Wert der jeweiligen Ausprägung einer Variablen in eckigen Klammern.

In einem weiteren Anonymisierungsschritt wurde die Anzahl der im SIAB-CF 7508 enthaltenen Variablen gegenüber dem SIAB-R 7508 reduziert. Insbesondere wurden die Erwerbshistorien vor Beginn des Eintritts in Arbeitslosigkeit vergrößert (vgl. Abschnitt 5).

Aus datenschutzrechtlichen Gründen musste weiterhin sichergestellt werden, dass die Informationen zu den Personen im SIAB-CF 7508 nicht direkt dem SIAB-R 7508 zugeordnet werden können. Für das SIAB-CF 7508 wurde daher auf Basis einer Zufallszahl eine neue Personennummer vergeben. Auch wurden für alle Personen die Beginn- und Enddaten der einzelnen Beobachtungen manipuliert. Dabei wurden nach dem Zufallsprinzip alle Beobachtungen einer Person zwischen 0 und 27 Tage nach vorne oder hinten verschoben.

## 5 Merkmalspektrum

Dieser Abschnitt liefert eine Übersicht zu den im SIAB-CF 7508 enthaltenen Merkmalen. Zunächst werden die Merkmale beschrieben, die sowohl im SIAB-CF 7508 als auch im SIAB-R 7508 enthalten sind. In Abschnitt 5.6 werden die Merkmale dargestellt, welche für das SIAB-CF 7508 neu gebildet wurden. Wie bereits im vorherigen Abschnitt dargelegt, beinhaltet das SIAB-CF 7508 nicht alle Merkmale, die im SIAB-R 7508 bereit gestellt werden. Diese Merkmale sind in Abschnitt 5.7 aufgeführt.

### 5.1 Identifikatoren

#### 5.1.1 Systemfreie Personennummer (persnr)

Variablenlabel	Systemfreie Personennummer
Variablenname	persnr
Kategorie	Identifikatoren
Herkunft	BeH, LeH, (X)LHG, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>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.</p> <p>Die systemfreie Personennummer aus dem SIAB-R 7508 wurde im SIAB-CF 7508 durch eine systemfreie, zufällig sortierte Personennummer ersetzt. Der Wertebereich dieser neu generierten Personennummer erstreckt sich von 1 bis 6.741.</p>



## 5.2 Gültigkeitszeitraum

### 5.2.1 Beginndatum Originalsatz (begorig)

Variablenlabel	Beginndatum Originalsatz
Variablenname	begorig
Kategorie	Gültigkeitszeitraum
Herkunft	BeH, LeH, (X)LHG, (X)ASU
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Das Beginndatum des Originalsatzes entspricht dem ursprünglichen Beginndatum der Meldung. Durch die Datumskorrekturen der Variable begorig im SIAB Regionalfiler kann diese auch im SIAB-CF 7508 von der Angabe in der SIAB abweichen. Die Variable begorig kann vom Beginndatum der Episoden (begepi) abweichen. Grund dafür ist das so genannte Episodensplitting (vgl. Dorner/König/Seth 2011, Abschnitt 3.5). In BeH-Meldungen sind Beginn- und Endjahr 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, (X)LHG, (X)ASU
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Das Enddatum des Originalsatzes entspricht dem ursprünglichen Enddatum der Meldung. Durch die Datumskorrekturen der Variable endorig im SIAB Regionalfiler kann diese auch im SIAB-CF 7508 von der Angabe in der SIAB abweichen. Die Variable endorig kann vom Enddatum der Episode (endepi) abweichen. Grund dafür ist das so genannte Episodensplitting (vgl. Dorner/König/Seth 2011, Abschnitt 3.5). In BeH-Meldungen sind Beginn- und Endjahr 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, (X)LHG, (X)ASU
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Diese Variable beschreibt das Beginndatum einer Episode. Mit dem Begriff „Episode“ wird der Zeitraum bezeichnet, den eine Datenzeile umfasst. Dieser Zeitraum kann durch Rechts- bzw. Linkszensierung einzelner Episoden (siehe Dorner/König/Seth 2011, Abschnitt 3.3.2) und/oder durch Episodensplitting vom Originalzeitraum der ursprünglichen Meldung abweichen (vgl. Dorner/König/Seth 2011, Abschnitt 3.5).

#### 5.2.4 Enddatum der Episode (endept)

Variablenlabel	Enddatum der Episode
Variablenname	endept
Kategorie	Generierter Gültigkeitszeitraum
Herkunft	BeH, LeH, (X)LHG, (X)ASU
Datentyp	Datum
Hierarchie	Keine
Detailbeschreibung	Diese Variable beschreibt das Enddatum einer Episode. Mit dem Begriff „Episode“ wird der Zeitraum bezeichnet, den eine Datenzeile umfasst. Dieser Zeitraum kann durch Rechts- bzw. Linkszensierung einzelner Episoden (vgl. Dorner/König/Seth 2011, Abschnitt 3.3.2) und/oder durch Episodensplitting vom Originalzeitraum der ursprünglichen Meldung abweichen (vgl. Dorner/König/Seth 2011, Abschnitt 3.5).

### 5.3 Informationen zur Person

#### 5.3.1 Geschlecht (frau)

Variablenlabel	Geschlecht
Variablenname	frau
Kategorie	Persönliche Merkmale
Herkunft	BeH, LeH, (X)LHG, (X)ASU
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 der SIAB 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, (X)LHG, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Das Geburtsjahr ist pro Konto konstant. Sätze mit einem Alter kleiner als 17 bzw. einem Alter größer als 62 Jahre wurden gelöscht (vgl. Dorner/König/Seth 2011, Abschnitt 3.3.3).
Qualitätshinweis	In den Originaldaten kann es vorkommen, dass das Geburtsjahr zwischen den Quellen wechselt. Dies wird bei der Datenaufbereitung der SIAB bereinigt. Hierbei kommt der Angabe aus der Sozialversicherungsnummer die höchste Priorität zu.

### 5.3.3 Ausbildung (bild)

Variablenlabel	Ausbildung
Variablenname	bild
Kategorie	Persönliche Merkmale
Herkunft	BeH, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Zu beachten ist, dass das Merkmal je nach Quelle eine unterschiedliche Bedeutung hat  <b>1) BeH</b>

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. Bundesagentur für Arbeit (Hrsg.) 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.

„Zusammen mit einem Betriebswechsel treten häufig auch Änderungen im Ausbildungsstatus auf. Das liegt daran, dass beim neuen Betrieb die Meldedaten neu zusammengestellt werden. Wenn ein Beschäftigter z.B. durch berufsbegleitende Weiterbildung einen höheren Abschluss erlangt hat, dann wird dieser Statuswechsel wahrscheinlich erst von einem neuen Beschäftigungsbetrieb erfasst. Es ist in der Regel davon auszugehen, dass bei längeren Beschäftigungszeiten innerhalb eines Betriebs die einmal erhobenen personenbezogenen Daten fortgeschrieben werden.“ (Meinken/Koch 2004, S. 63)

Eine Methode, fehlende Werte oder inkonsistente zeitliche Verläufe der Bildungsangaben in der Vorgängerstichprobe IABS zu bereinigen, findet sich in Fitzenberger/Osikominu/Völter (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.

## 2) (X)ASU

	<p>Für (X)ASU-Sätze ist die abgeschlossene Berufsausbildung zum Beginn der Arbeitsuche enthalten. Folgende Ausprägungen kommen vor:</p> <p>21 ohne abgeschlossene Berufsausbildung  22 betriebliche Ausbildung  23 außerbetriebliche Ausbildung  24 Berufsfachschule  25 Fachschule  26 Fachhochschule  27 Universität</p>
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 Beschäftigten).“ (Meinken/Koch 2004, S. 63).

#### 5.3.4 Schulausbildung (schbild)

Variablenlabel	Schulausbildung
Variablenname	schbild
Kategorie	Persönliche Merkmale
Herkunft	(X)ASU
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 kontinuierlich bis 2008, bleibt aber problematisch. Die Quellen ASU und XASU können allerdings im faktisch anonymisierten Datensatz nicht unterschieden werden, und somit auch nicht im CF.

### 5.4 Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche

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

Variablenlabel	Abgabegrund/ Beendigungsgrund/ SGB II-Einstellungsgrund/ Abmeldegrund
Variablenname	grund_gr
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, (X)LHG, (X)ASU
Datentyp	Numerisch

Hierarchie	Keine
Detailbeschreibung	<p><b>1) BeH</b>  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 (Jahres-, Unterbrechungs- und Abmeldungen), während Anmeldungen mit fehlenden 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 1.1.1999 geltenden Meldeverfahrens (gemäß DEÜV) kodiert.</p> <p><b>2) LeH</b>  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.</p> <p><b>3) (X)LHG</b>  Bei (X)LHG Datensätzen ist der SGB II-Einstellungsgrund enthalten und gibt den Grund für die Beendigung von laufenden Leistungen an. Der SGB II-Beendigungsgrund bezieht sich auf die Person, nicht aber auf die Bedarfsgemeinschaft. Das Merkmal ist genau zum Satzende 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 SGB II-Leistungsbezug beendet, werden alle Sätze der Mitglieder in einer Bedarfsgemeinschaft an diesem Datum gesplittet, der SGB II-Beendigungs-/Einstellungsgrund ist jedoch nur für die betroffene Person gefüllt.</p> <p><b>4) (X)ASU</b></p>

	<p>Bei (X)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.</p> <p>Im Zuge der Anonymisierung der SIAB wurden Abgabegründe innerhalb der vier im SIAB Regionalfile enthaltenen Quellen zusammengefasst.</p>
Qualitätshinweise	Für den Anteil valider Werte (Füllgrad) des Abgabegrundes in den Quellen (X)LHG und (X)ASU vgl. Dorner/König/Seth (2011), S. 37 und 38.

#### 5.4.2 Tagesentgelt, täglicher Leistungssatz (tentgelt\_gr)

Variablenlabel	Tagesentgelt, täglicher Leistungssatz
Variablenname	tentgelt_gr
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p><b>1) BeH</b></p> <p>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 der IEB in Kalendertagen. Das Tagesentgelt ist für den gesamten Zeitraum einheitlich in der Währung Euro angegeben. Im SIAB Regionalfile wird das Tagesentgelt aus datenschutzrechtlichen Gründen nur gerundet in vollen Euro Werten ausgewiesen. Entgelte unter 0,5 Euro wurden dabei auf 0 Euro abgerundet.</p> <p>Bis 1998 meldeten Arbeitgeber grundsätzlich nur das sozialversicherungspflichtige Arbeitsentgelt. Entgelte, die unterhalb der Geringfügigkeitsgrenze lagen, wurden nicht gemeldet.</p> <p>Entgelte, welche die Beitragsbemessungsgrenze in der gesetzlichen Rentenversicherung überschritten, werden 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. Eine Unterscheidung welcher Rentenversicherungsträger für eine Person zuständig ist, fehlt in den Daten.</p>

Seit der Einbeziehung der geringfügig Beschäftigten in das Meldeverfahren zum 01.04.1999 werden auch Entgelte unterhalb der Geringfügigkeitsgrenze erfasst, die Beitragsbemessungsgrenze als obere Kappungsgrenze gilt jedoch weiterhin. In einigen Fällen übersteigen die gemeldeten Entgelte die Beitragsbemessungsgrenze allerdings trotzdem. Dies dürfte in der Regel auf Jahressonderzahlungen zurückzuführen sein, die der Arbeitgeber bei den Jahres-, Unterbrechungs- oder Abmeldungen zum normalen Entgelt hinzurechnen kann. Dabei ist es dann unerheblich, ob durch diese Addition die für den Meldezeitraum maßgebliche Beitragsbemessungsgrenze in der Rentenversicherung überschritten wird. Es könnte sich allerdings auch um fehlerhafte Angaben bei Beschäftigungszeit oder Entgelthöhe handeln, was wegen der Versicherungsrelevanz dieser Informationen jedoch eher selten der Fall sein dürfte.

Aus datenschutzrechtlichen Gründen werden alle Entgelte, die oberhalb der Beitragsbemessungsgrenze (in vollen Euro) liegen, nicht mit ihrem tatsächlichen Wert ausgewiesen sondern durch den im Jahr der Meldung gültigen Euro-Betrag der Beitragsbemessungsgrenze ersetzt. 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> im Menüunterpunkt "Arbeitshilfen" zur SIAB.

Eine Angabe von 0 Euro ist neben den o.g. abgerundeten Entgelten unterhalb von 0,5 Euro auch 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.

## **2) LeH**

Bei LeH-Datensätzen enthält das Merkmal den täglichen Leistungssatz, jeweils umgerechnet und gerundet auf volle Euro. Leistungssätze unter 0,5 Euro wurden dabei auf 0 Euro abgerundet.

Es ist zu berücksichtigen, dass der tägliche Leistungssatz bei Datensätzen mit einem Original-Beginndatum, das vor dem 1.1.1998 liegt, für Werktage und bei Datensätzen mit einem Original-Beginndatum ab 1.1.1998 für Kalendertage gilt.

Die Angabe von 0 Euro kann entweder durch die Abrundung sehr geringer Leistungssätze unterhalb von 0,5 Euro verursacht werden oder seit dem 1.1.2005 auch im Zusammenhang mit Sperrzeiten oder Leistungsunterbrechungen stehen. 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 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	<p>Die berufliche Stellung des Beschäftigten im Meldezeitraum wird vom Arbeitgeber im Rahmen der „Angaben zur Tätigkeit“ übermittelt.</p> <p>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.</p> <p>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).</p>



Falls mehrere Schlüsselzahlen in Betracht kommen, ist der Arbeitgeber dazu angehalten, die Einstufung nach der überwiegend ausgeübten Tätigkeit vorzunehmen. Lässt sich diese nicht eindeutig feststellen, ist die Schlüsselzahl der höherwertigen Stellung im Beruf einzutragen (vgl. Bundesagentur für Arbeit (Hrsg.) 2005, S. VI).

#### 5.4.4 Erwerbsstatus (erwstat\_gr)

Variablenlabel	Erwerbsstatus vergrößert
Variablenname	erwstat_gr
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, LeH, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal nimmt für jede Datenquelle unterschiedliche Werte mit unterschiedlichen Bedeutungen an.</p> <p><b>1) BeH</b> Für BeH-Sätze ist die Personengruppe der Beschäftigtenmeldung enthalten, die mit dem neuen Meldeverfahren zum 1.1.1999 eingeführt wurde. Sie weist beitrags- oder leistungsrechtliche Besonderheiten des Beschäftigungsverhältnisses aus, die mindestens einen Sozialversicherungsträger betreffen, unterschieden nach meldenden Institutionen.</p> <p>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.</p> <p>Das Merkmal Personengruppe ist erst seit der Einführung des neuen Meldeverfahrens zum 1.1.1999 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.</p>

## 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) (X)ASU

Bei (X)ASU ist der Status der Arbeitssuche enthalten. Eine Unterscheidung nach der Herkunft aus den Originalquellen ASU und XASU ist durch deren Zusammenfassung nicht mehr möglich. Aus der XASU wurden nur die Statusangaben „nicht arbeitslos, aber arbeitssuchend“ bzw. „arbeitssuchend und gleichzeitig arbeitslos“ in die SIAB integriert. Diese Angaben zum Status der Arbeitssuche wurden mit dem entsprechenden Status aus der ASU zu einer gemeinsamen Kategorie zusammengefasst.

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.

## 5.5 Generierte technische Merkmale

### 5.5.1 Quelle des Satzes (quelle\_gr)

Variablenlabel	Quelle des Satzes, vergrößert
Variablenname	quelle_gr
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, (X)LHG, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Anonymisierung	Keine
Detailbeschreibung	Das Merkmal gibt die Datenquelle an. Dabei wurden die Originalquellen ASU und XASU sowie LHG und XLHG jeweils zu einer Quelle zusammengefasst. Eine Unterscheidung nach der Trägerart ist aus Datenschutzgründen somit nicht mehr möglich.

Durch Korrekturen unterscheiden sich die ausgewiesenen Zeiträume der Quellen im Vergleich zur SIAB 7508. Informationen zur Aufbereitung entnehmen Sie bitte Dorner/König/Seth (2011), Kapitel 3.

### 5.5.2 Satzzähler pro Episode (level2)

Variablenlabel	Satzzähler pro Episode
Variablenname	level2
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, (X)LHG, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Anonymisierung	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, (X)ASU, (X)LHG). 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.

### 5.5.3 Satzzähler pro Episode und Quelle (level1)

Variablenlabel	Satzzähler pro Episode und Quelle
Variablenname	level1
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, (X)LHG, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Anonymisierung	Keine

Detailbeschreibung	<p>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 parallele Sätze zählt, zusätzlich auch noch nach der Quelle dieser Sätze. Die Level1-Variable nimmt somit 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.</p> <p>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.)</p>
--------------------	---

## 5.6 Neu gebildete Merkmale im SIAB-CF 7508

Für die Zeiten in sozialversicherungspflichtiger Beschäftigung, geringfügiger Beschäftigung und Leistungsbezug vor dem ersten Bezug von Arbeitslosengeld I im Jahr 2005 beinhaltet das SIAB-CF 7508 nur vergrößerte Informationen. Weiterhin beinhaltet das SIAB-CF 7508 Angaben zum Arbeitsort nur in extrem vergrößerter Form. Im SIAB-CF 7508 lässt sich lediglich unterscheiden, ob eine Person in Ost- oder Westdeutschland beschäftigt ist. Ein weiteres Merkmal, welches für die SIAB-CF 7508 generiert wurde, gibt an, ob der letzten Arbeitsort einer Person vor Eintritt in den Bezug von Arbeitslosengeld I im Jahr 2005 in Ost- oder Westdeutschland war.

### 5.6.1 Tage in soz.pfl. Besch. in den 10 Jahren vor Arbeitslosengeldbezug (tage\_svb\_gr)

Variablenlabel	Tage in soz.pfl. Besch. in den 10 Jahren vor Arbeitslosengeldbezug
Variablenname	tage_svb_gr
Kategorie	Neues Merkmal SIAB-CF 7508
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Anonymisierung	Keine

Detailbeschreibung	Dieses Merkmal gibt die Anzahl der Tage an, an denen eine Person in den zehn Jahren vor dem Beginn des ersten Bezugs von Arbeitslosengeld im Jahr 2005 sozialversicherungspflichtig beschäftigt war. Da es in Hinblick auf den Datenschutz als kritisch erscheint, dass für jede Person die kumulierte Zahl der Tage in sozialversicherungspflichtiger Beschäftigung für die vorangegangenen zehn Jahre enthalten ist, wird dieses Merkmal auf Basis der Perzentile ihrer Verteilung gruppiert und ist im SIAB-CF 7508 in dieser vergrößerten Form enthalten. Diese Information wurde auf alle Spells einer Person übertragen.
--------------------	--

### 5.6.2 Tage in geringf. Besch. in den 10 Jahren vor Arbeitslosengeldbezug (tage\_grfb\_gr)

Variablenlabel	Tage in geringf. Besch. in den 10 Jahren vor Arbeitslosengeldbezug
Variablenname	tage_grfb_gr
Kategorie	Neues Merkmal SIAB-CF 7508
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Anonymisierung	Keine
Detailbeschreibung	Dieses Merkmal gibt die Anzahl der Tage an, an denen eine Person in den zehn Jahren vor dem Beginn des ersten Bezuges von Arbeitslosengeld im Jahr 2005 geringfügig beschäftigt war. Da es in Hinblick auf den Datenschutz als kritisch erscheint, dass für jede Person die kumulierte Zahl der Tage in geringfügiger Beschäftigung für die vorangegangenen zehn Jahre enthalten ist, wird dieses Merkmal auf Basis der Perzentile ihrer Verteilung gruppiert und ist im SIAB-CF 7508 in dieser vergrößerten Form enthalten. Diese Information wurde auf alle Spells einer Person übertragen.

### 5.6.3 Tage in Leistungsbezug in den 10 Jahren vor Arbeitslosengeldbezug (tage\_leist\_gr)

Variablenlabel	Tage in Leistungsbezug in den 10 Jahren vor Arbeitslosengeldbezug
Variablenname	tage_leist_gr
Kategorie	Neues Merkmal SIAB-CF 7508
Herkunft	LeH
Datentyp	Numerisch
Hierarchie	Keine

Anonymisierung	Keine
Detailbeschreibung	Dieses Merkmal gibt die Anzahl der Tage an, an denen eine Person in den zehn Jahren vor dem Beginn des ersten Bezugs von Arbeitslosengeld im Jahr 2005 Leistungen nach dem Rechtskreis SGB III (Arbeitslosengeld, Arbeitslosenhilfe oder Unterhaltsgeld) bezogen hat. Da es in Hinblick auf den Datenschutz als kritisch erscheint, dass für jede Person die kumulierte Zahl der Tage im Leistungsbezug für die vorangegangenen zehn Jahre enthalten ist, wird dieses Merkmal auf Basis der Perzentile ihrer Verteilung gruppiert und ist im SIAB-CF 7508 in dieser vergrößerten Form enthalten. Diese Information wurde auf alle Spells einer Person übertragen.

#### 5.6.4 Arbeitsort Ost/West (ao\_ost)

Variablenlabel	Arbeitsort Ost/West
Variablenname	ao_ost
Kategorie	Neues Merkmal SIAB-CF 7508
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Anonymisierung	Keine
Detailbeschreibung	Dummy für den Arbeitsort einer Person (0 - Westdeutschland, 1 - Ostdeutschland). Dieses Merkmal gibt an, ob der Betrieb in West- oder Ostdeutschland abgesiedelt ist, in dem eine Person beschäftigt ist. Berlin wird Ostdeutschland zugeordnet. Dieses Merkmal ist nur für Meldungen aus der BeH gefüllt.

#### 5.6.5 Arbeitsort letzte Besch. vor Arbeitslosengeldbezug Ost/West (letzte\_besch\_ost)

Variablenlabel	Arbeitsort letzte Besch. vor Arbeitslosengeldbezug Ost/West
Variablenname	letzte_besch_ost
Kategorie	Neues Merkmal SIAB-CF 7508
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Anonymisierung	Keine

Detailbeschreibung	Dummy für den Arbeitsort der letzten Beschäftigung einer Person vor dem ersten Bezug von Arbeitslosengeld im Jahr 2005 (0 - Westdeutschland, 1 - Ostdeutschland). Dieses Merkmal gibt an, ob der Betrieb in West- oder Ostdeutschland angesiedelt ist, in dem eine Person vor Eintritt in den Bezug von Arbeitslosengeld I im Jahr 2005 zuletzt beschäftigt war. Berlin wird Ostdeutschland zugeordnet. Diese Information wurde auf alle Spells einer Person übertragen.
--------------------	--

## 5.7 Zusätzliche Merkmale im SIAB-R 7508

Dieser Abschnitt liefert einen Überblick zu den Merkmalen, die im SIAB-R 7508 enthalten sind aber nicht in das SIAB-CF 7508 aufgenommen wurden.

### 5.7.1 Betriebsnummernzähler (bnn)

Variablenlabel	Betriebsnummernzähler
Variablenname	bnn
Kategorie	Identifikatoren
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Der Betriebsnummernzähler nummeriert die Betriebe im Erwerbsleben einer Person im Zeitverlauf aufsteigend.</p> <p>Beispiel: Der erste Betrieb, in dem eine Person beschäftigt war, erhält den Wert 1. Wechselt die Person zu einem anderen Betrieb, erhält dieser den Wert 2. Für nachfolgende Betriebe erhöht sich der Wert um jeweils 1. Wechselt die Person zu einem Betrieb, in welchem sie zu einem früheren Zeitpunkt bereits beschäftigt war, so erhält dieser Betrieb den Wert, der für die erste Beschäftigung galt (z.B. 2). Wechselt eine Person nach genau einem Wechsel zum ersten Betrieb zurück, ergäbe sich daher im Zeitverlauf für die Variable bnn die Abfolge 1-2-1.</p> <p>Fehlende Betriebsnummern in den Originaldaten wurden auch im SIAB Regionalfile mit missing (.z) gekennzeichnet.</p>

### 5.7.2 Staatsangehörigkeit (deutsch)

Variablenlabel	Staatsangehörigkeit deutsch
Variablenname	deutsch
Kategorie	Persönliche Merkmale

Herkunft	BeH, LeH, (X)LHG, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Das Merkmal unterscheidet für Personen, deren Arbeitsort in Westdeutschland liegt, zwischen deutscher und nicht-deutscher Staatsangehörigkeit.</p> <p>Für Personen, die in ihrem Erwerbsverlauf mindestens einmal in Ostdeutschland beschäftigt waren, ist keine Angabe zur Staatsangehörigkeit enthalten.</p> <p>Ebenfalls nicht ausgewiesen wird die Staatsangehörigkeit bei Personen, für die in keiner Meldung ein gültiger Arbeitsort vorliegt und somit die Zuordnung zu Ost- bzw. Westdeutschland nicht möglich ist.</p>

### 5.7.3 Beruf – ausgeübte Tätigkeit (beruf\_gr)

Variablenlabel	Beruf - ausgeübte Tätigkeit, vergrößert
Variablenname	beruf_gr
Kategorie	Informationen zu Beschäftigung, Leistungsbezug und Arbeitsuche
Herkunft	BeH, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p><b>1) BeH</b></p> <p>Die Berufsbezeichnung der im Meldezeitraum vom Beschäftigten ausgeübten Tätigkeit ist Bestandteil der vom Arbeitgeber übermittelten „Angaben zur Tätigkeit“.</p> <p>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. Bundesagentur für Arbeit (Hrsg.) 2005, S. V).</p> <p>Im Rahmen der Anonymisierung des SIAB Regionalfiler werden die aus der BeH stammenden Berufsangaben zu 120 Berufsgregaten zusammengefasst.</p>



	<p><b>2) (X)ASU</b>  Bei der Anonymisierung wurde diese Variable für (X)ASU-Datensätze auf missing (.z) gesetzt.</p>
--	--

#### 5.7.4 Beschäftigungsdauer vor dem Alter von 17 (tage\_jung)

Variablenlabel	Tage in Beschäftigung vor der Altersuntergrenze (Jahr des 17. Geburtstags)
Variablenname	tage_jung
Kategorie	Informationen zu Beschäftigung
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Dieses Merkmal gibt an, wie viele Tage eine Person vor dem Jahr ihres 17. Geburtstags beschäftigt war. Gezählt wird dabei nur die Hauptbeschäftigung (Stata Syntax: quelle_gr==1 & level2==0).

#### 5.7.5 Beschäftigungsdauer nach dem Alter von 62 (tage\_alt)

Variablenlabel	Tage in Beschäftigung nach der Altersobergrenze (Jahr des 62. Geburtstags)
Variablenname	tage_alt
Kategorie	Informationen zu Beschäftigung
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	Dieses Merkmal gibt an, wie viele Tage eine Person nach dem Jahr ihres 62. Geburtstags beschäftigt war. Gezählt wird dabei nur die Hauptbeschäftigung (Stata Syntax: quelle_gr==1 & level2==0).

#### 5.7.6 Wirtschaftszweig 73 – Wirtschaftsgruppe 73 (w73\_gen\_gr)

Variablenlabel	Wirtschaftszweig 73, vergrößert, ab 2003 generiert
Variablenname	w73_gen_gr
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine

Detailbeschreibung	<p>Dieses Merkmal basiert auf der Wirtschaftszweigklassifikation WS73. 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.</p> <p>Jedem Betrieb darf nur eine Kennziffer zugeordnet werden. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, muss die wirtschaftliche Haupttätigkeit ermittelt werden.</p> <p>Angaben zur WS73 liegen nur von 1975 bis 2002 vor. Die Variable enthält für diesen Zeitraum die Angaben aus den Meldungen basierend auf der Wirtschaftsklasse W73. Ab 2003 werden sämtliche Werte, basierend auf einer Heuristik, ergänzt. Dazu werden zunächst gültige Werte in die Zukunft fortgeschrieben. Ist eine Fortschreibung nicht möglich, so werden weitere fehlende Werte anhand von Umschlüsselungstabellen ersetzt. Zur Methodik und für Qualitätshinweise sei auf Eberle u. a. (2011) verwiesen.</p> <p>Im Rahmen der Anonymisierung des SIAB Regionalfiles werden die Wirtschaftszweigangaben zu 16 Aggregaten zusammengefasst.</p> <p>In einzelnen Regionen wurden Wirtschaftszweige, die in mindestens einem der geprüften Jahre unter der Mindestfallzahl für die faktische Anonymität lagen, anonymisiert und auf missing (.z) umgesetzt (vgl. Dorner/König/Seth 2011, Anhang)</p>
--------------------	--

### 5.7.7 Wirtschaftszweig 03 – Wirtschaftsgruppe 03 (w03\_gr)

Variablenlabel	Wirtschaftszweig 03, vergrößert
Variablenname	w03_gr
Kategorie	Betriebsmerkmale
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>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 seit 2003 gefüllt.</p>

Jedem Betrieb darf nur eine Kennziffer zugeordnet werden. Ist ein Betrieb in unterschiedlichen wirtschaftlichen Bereichen tätig, muss die wirtschaftliche Haupttätigkeit ermittelt werden.

In einzelnen Regionen wurden Wirtschaftszweige, die in mindestens einem der geprüften Jahre unter der Mindestfallzahl für die faktische Anonymität lagen, anonymisiert und auf missing (.z) umgesetzt (vgl. Dorner/König/Seth 2011, Anhang).

### 5.7.8 Arbeitsort Region (ao\_region)

Variablenlabel	Arbeitsort Region
Variablenname	ao_region
Kategorie	Ortsangaben
Herkunft	BeH
Datentyp	Numerisch
Hierarchie	Kreisregion / Bundesland
Detailbeschreibung	<p>Kreisfreie Städte und Landkreise wurden zu 333 Kreisregionen (vgl. Dorner/König/Seth 2011, Abschnitt 7.5) aggregiert, sodass jede dieser Regionen (zum 31.12.2008) eine Mindesteinwohnerzahl von 100.000 Einwohnern aufweist.</p> <p>Die Ausgangsdaten enthalten den Kreis (kreisfreie Stadt oder Landkreis), in dem der Betrieb des Beschäftigten angesiedelt ist. Der 5-stellige Kreisschlüssel enthält in den ersten beiden Stellen den Schlüssel des Bundeslands, in der ersten bis dritten Stelle den Regierungsbezirk und in der ersten bis fünften Stelle den Kreis. Bei Ländern ohne Regierungsbezirk steht an der dritten Stelle eine Null.</p> <p>Um konsistente regionale Zuordnungen über den gesamten Beobachtungszeitraum sicherzustellen, wurden die zugrunde liegenden 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.</p>

### 5.7.9 Pendlerstatus (pendler)

Variablenlabel	Pendlerstatus
Variablenname	pendler

Kategorie	Ortsangaben
Herkunft	BeH, LeH, (X)LHG, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Detailbeschreibung	<p>Dieses Merkmal gibt an, ob der aktuelle oder letzte Arbeitsort einer Person dem aktuellen Wohnort entspricht.</p> <p>Als Referenz für alle Quellen wird der aktuelle bzw. der Arbeitsort der Person aus der letzten zeitlich vorher liegenden BEH-Meldung (Hauptbeschäftigung) zugrunde gelegt und mit dem aktuellen Wohnort verglichen.</p> <p>Das Merkmal weist folgende Ausprägungen aus:</p> <p>0: Gemeinde des Wohnortes entspricht der Gemeinde des (letzten gefüllten) Arbeitsortes (kein Pendler)</p> <p>1: Gemeinde des Wohnortes entspricht nicht der Gemeinde des (letzten gefüllten) Arbeitsortes, beide Gemeinden gehören aber der gleichen Kreisregion an (intra-regionaler Pendler).</p> <p>2: Gemeinde des Wohnortes und Gemeinde des (letzten gefüllten) Arbeitsortes liegen in unterschiedlichen Kreisregionen (inter-regionaler Pendler)</p> <p>Das Merkmal ist gefüllt für:</p> <p>BEH: 1999-2008</p> <p>LEH: 2003-2008</p> <p>(X)ASU: 2000-2004 und 2007-2008</p> <p>(X)LHG: 2007-2008</p> <p>Ist für eine Episode einer Person in den Originaldaten der Wohnort einer Quelle nicht gefüllt, so wird das Merkmal Pendlerstatus auf .z gesetzt. Ebenso enthält das Merkmal .z, wenn zwar der Wohnort der Person in einer Quelle gefüllt ist, jedoch im Erwerbsverlauf vorher noch kein BEH-Satz mit gefülltem Arbeitsort als Referenzort existiert.</p>

#### 5.7.10 Satzzähler Konto (spell)

Variablenlabel	Satzzähler Konto
Variablenname	spell
Kategorie	Generierte technische Merkmale
Herkunft	BeH, LeH, (X)LHG, (X)ASU
Datentyp	Numerisch
Hierarchie	Keine
Anonymisierung	Keine

Detailbeschreibung	Der Satzzähler Konto nummeriert aufsteigend die Sätze einer Person. Das Merkmal wird beim Episodensplitting erzeugt und bezieht sich auf die gesplitteten Sätze. Mit dem Merkmal Satzzähler Konto kann die ursprüngliche Sortierung auf einfache Art und Weise wieder hergestellt werden. Die Sortierung erfolgt grundsätzlich zunächst nach dem Beginndatum der Episode und dann nach der Quelle.
--------------------	--

## 6 Arbeiten mit dem Campus File

Nach dem PASS Campus File (PASS-CF), welches auf dem Panel "Arbeitsmarkt und soziale Sicherung" (PASS) beruht, bietet das FDZ mit dem SIAB-CF 7508 ein zweites Datenprodukt dieser Art an. Ebenso wie das PASS-CF, welches in Kirchner/Gschwind (2011) dokumentiert ist, wurde auch das SIAB-CF 7508 speziell für Lehrzwecke entwickelt und soll der anschaulichen Vermittlung von Erhebungs-, Datenmanagement- und Analysetechniken anhand verschiedener ökonomischer und sozialwissenschaftlicher Fragestellungen dienen. Grundsätzlich sind diese Daten keinesfalls für inhaltliche Auswertungen irgendeiner Art oder Publikationen, also beispielsweise auch nicht zur Verfassung von Seminar-, Bachelor- oder Masterarbeiten geeignet. Es ist nicht möglich mit diesen Daten generalisierbare Aussagen über einzelne Merkmale und Zusammenhänge zwischen verschiedenen Merkmalen zu treffen.

In Abgrenzung zum PASS-CF beruht das SIAB-CF 7508 ausschließlich auf administrativer Datenbasis. Es sind keinerlei Befragungsdaten enthalten. Das SIAB-CF 7508 eignet sich insbesondere für das Vermitteln von Arbeiten mit Datensätzen im Spell-Format. Seine Struktur bildet detaillierte Erwerbsbiographien ab und lässt die Anwendung von Methoden wie der Verweildaueranalyse zu. Des Weiteren lassen sich im SIAB-CF 7508 Zugänge in und Abgänge aus Maßnahmen der Förderung der beruflichen Weiterbildung (FbW) und Deutschsprachlehrgängen (DSL) identifizieren. Dies erlaubt auch die Vermittlung von Methoden mikroökonomischer Wirkungsanalysen mittels des SIAB-CF 7508.

## 7 Abschließende Hinweise und Datenzugang

Campus Files sind absolut anonymisierte Datensätze, die vom IAB speziell für die Nutzung in der Lehre an Universitäten, Hochschulen und Forschungsinstituten entwickelt wurden. Aufgrund der umfangreichen Anonymisierungsverfahren, die bei der Generierung der CF angewandt wurden, dienen die Daten ausschließlich der Vermittlung von Erhebungs-, Datenmanagement- und Analysetechniken. Für valide inhaltliche Analysen sind die CF nicht geeignet. Hierfür sollten die entsprechenden Scientific Use Files oder schwach anonymisierten Versionen der zugrundeliegenden Datensätze verwendet werden.

Das SIAB-CF 7508 wird vom FDZ als kostenloser Download angeboten. Nach einer Registrierung und Zustimmung zu den Nutzungsbedingungen können die Daten mit einem

Passwort abgerufen werden. Eine Weitergabe der Daten im Rahmen von Lehrveranstaltungen ist möglich, es ist dabei jedoch ausdrücklich auf die Nutzungsbedingungen und Einschränkungen im Arbeiten mit diesen hinzuweisen. Die kommerzielle Nutzung sowie eine kommerzielle Weitergabe sind untersagt.

Weitere Informationen zu den Campus Files finden sich auf der FDZ-Homepage [http://fdz.iab.de/de/Campus\\_Files.aspx](http://fdz.iab.de/de/Campus_Files.aspx).

## Literatur

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

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

Dorner, Matthias; König, Marion; Seth, Stefan (2011): Stichprobe der Integrierten Arbeitsmarktbiografien. Regionalfile 1975-2008 (SIAB-R 7508). FDZ-Datenreport, 07/2011, Nürnberg, 74 S.

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

Eberle, Johanna; Jacobebbinghaus, Peter; Ludsteck, Johannes; Witter, Julia (2011): Generation of time-consistent industry codes in the face of classification changes. Simple heuristic based on the Establishment History Panel (BHP). FDZ-Methodenreport, 05/2011 (en), Nürnberg, 21 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, Bd. 126, Nr. 3, S. 405–436.

Hochfellner, Daniela; Müller, Dana; Schmucker, Alexandra; Roß, Elisabeth (2012): Datenschutz am Forschungsdatenzentrum. FDZ-Methodenreport, 06/2012 (de), Nürnberg, 27 S.

Kirchner, Antje; Gschwind, Lutz (2011): Panel "Arbeitsmarkt und soziale Sicherung" - Die PASS Campus Files \* Datensätze für den Einsatz in der Lehre. FDZ-Datenreport, 06/2011, Nürnberg, 22 S.

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

Ronning, Gerd; Sturm, Roland; Höhne, Jörg; Lenz, Rainer; Rosemann, Martin; Scheffler, Michael; Vorgrimler, Daniel (2005): Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten, Bd. 4 von Statistik und Wissenschaft. Wiesbaden: Statistisches Bundesamt.

Schermer, Joachim (1999): Das Meldeverfahren zur Sozialversicherung nach der DEÜV ab 1. 1. 1999. Meldungen nach den Vorschriften des Sozialversicherungsrechts. Frechen-Königsdorf: Datakontext-Fachverl., 2 Aufl..

Südfeld, Erwin (1987): Anonymisierungsstandards und generelle Abwicklungsregeln für Anforderungen nach anonymisierten Einzelangaben im Statistischen Bundesamt. In: Bundesamt, Statistisches (Hrsg.) Nutzung von anonymisierten Einzelangaben aus Daten der amtlichen Statistik, Bd. 5 von Forum der Bundesstatistik, Stuttgart: Kohlhammer, S. 146–156.

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, Bd. 21, Nr. 4, S. 468–482.

Zwick, Markus (2008): CAMPUS Files. Kostenfreie Public Use Files für die Lehre. In: AStA Wirtschafts- und Sozialstatistisches Archiv, Bd. 2, Nr. 1–2, S. 175–188.



## Impressum

### FDZ-Datenreport 05/2013

#### Herausgeber

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

#### Redaktion

Stefan Bender, Heiner Frank

#### Technische Herstellung

Heiner Frank

#### Rechte

Nachdruck - auch auszugsweise - nur mit  
Genehmigung des FDZ gestattet

#### Bezugsmöglichkeit

[http://doku.iab.de/fdz/reporte/2013/DR\\_05-13.pdf](http://doku.iab.de/fdz/reporte/2013/DR_05-13.pdf)

#### Internet

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

#### Rückfragen zum Inhalt an:

Matthias Umkehrer  
Institut für Arbeitsmarkt- und  
Berufsforschung (IAB)  
Regensburger Str. 104  
90478 Nürnberg  
Telefon: 0911 / 179-6211  
<mailto:matthias.umkehrer@iab.de>