

Research Data Centre (FDZ)  
of the German Federal  
Employment Agency (BA)  
at the Institute for  
Employment Research (IAB)

FDZ

# FDZ-Datenreport

Documentation of labour market data

09/2013

EN

## Sample-of-Integrated-Labour- Market-Biographies Regional-File 1975-2010 (SIAB-R 7510)

Philipp vom Berge  
Anja Burghardt  
Simon Trenkle



Bundesagentur für Arbeit

# Sample of Integrated Labour Market Biographies – Regional File 1975-2010 (SIAB-R 7510)

Philipp vom Berge, Anja Burghardt, Simon Trenkle  
(Institute for Employment Research, IAB)

Documentation: Version 1 (SIAB\_Regional\_7510\_EN\_v1\_dok1)

Data: Version 1 (SIAB Regional 7510 v1)

Die FDZ-Datenreporte beschreiben die Daten des FDZ im Detail. Diese Reihe hat somit eine doppelte Funktion: zum einen stellen Nutzerinnen und Nutzer fest, ob die angebotenen Daten für das Forschungsvorhaben geeignet sind, zum anderen dienen sie zur Vorbereitung der Auswertungen.

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of reports has a dual function: on the one hand, those using the reports can ascertain whether the data offered is suitable for their research task; on the other, the data can be used to prepare evaluations.

## Content

<b>1</b>	<b>INTRODUCTION AND OUTLINE</b> .....	<b>8</b>
1.1	Introduction .....	8
1.1	Outline .....	9
1.2	List of variables .....	10
1.2	Volume Structure.....	12
1.3	Changes as compared to SIAB Regional 7508 .....	12
1.4	Data use .....	13
<b>2</b>	<b>DATA SOURCES</b> .....	<b>13</b>
2.1	Employee History (BeH).....	14
2.1	Benefit Recipient History (LeH) .....	15
2.2	Unemployment Benefit II Recipient History (LHG / XLHG) .....	15
2.3	Jobseeker Histories (ASU/XASU).....	17
<b>3</b>	<b>DATA PREPARATION AND SAMPLING PROCEDURE</b> .....	<b>17</b>
3.1	Corrections and validation procedures of the original data for SIAB-R 7510.....	17
3.1.1	Employee History (BeH).....	17
3.1.2	Benefit Recipient History (LeH).....	18
3.1.3	Unemployment Benefit II Recipient History (LHG / XLHG).....	18
3.1.4	Jobseeker History (ASU / XASU).....	18
3.1.5	SGB II anonymisation.....	19
3.2	Episode-Splitting .....	20
3.3	Sampling procedure.....	21
3.4	Anonymisation of the SIAB-R 7510.....	21
3.4.1	Merging sources .....	22
3.4.2	Periods of validity of the sources .....	22
3.4.3	Censoring of data on age.....	25
3.4.4	Aggregation and deletion of variables.....	25
3.4.5	Consequences of the anonymisation .....	27
3.5	Missing values .....	28
<b>4</b>	<b>DATA QUALITY AND PROBLEMS</b> .....	<b>28</b>
4.1	Employee History (BeH).....	30
4.2	Benefit Recipient History (LeH) .....	31

<b>4.3</b>	<b>Unemployment Benefit II Recipient History (LHG/XLHG)</b> .....	<b>31</b>
<b>4.4</b>	<b>Jobseeker History (ASU/XASU)</b> .....	<b>32</b>
4.4.1	ASU .....	32
4.4.2	XASU .....	33
<b>5</b>	<b>DESCRIPTION OF VARIABLES</b> .....	<b>33</b>
<b>5.1</b>	<b>Identifiers</b> .....	<b>33</b>
5.1.1	Systemfreie Personennummer (persnr).....	33
5.1.2	Establishment counter (bnn).....	34
<b>5.2</b>	<b>Period of validity</b> .....	<b>34</b>
5.2.1	Original start date of observation (begorig).....	34
5.2.2	Original end date of observation (endorig).....	35
5.2.3	Start date of split episode (begepi).....	35
5.2.4	Start date of split episode (endeipi).....	35
<b>5.3</b>	<b>Personal information</b> .....	<b>35</b>
5.3.1	Gender (frau) .....	35
5.3.2	Year of birth (gebjahr).....	36
5.3.3	Nationality (nation) .....	36
5.3.4	School education and vocational training (bild).....	36
5.3.5	School-leaving qualification (schbild) .....	38
<b>5.4</b>	<b>Information on employment, benefit receipt and job-search</b> .....	<b>38</b>
5.4.1	Reason for notification/ reason for end of benefit receipt/ reason for discontinuation of unemployment benefit II/ reason for deregistration, aggregated (grund_gr).....	38
5.4.2	Daily wage, daily benefit rate, aggregated (tengtelt_gr).....	40
5.4.3	Beruf – ausgeübte bzw. letzte Tätigkeit (beruf_gr) .....	41
5.4.4	Occupational status and working hours (stib) .....	42
5.4.5	Employment status (erwstat_gr) .....	43
5.4.6	Employment before the age of 17 (tage_jung) .....	44
5.4.7	Employment after the age of 62 (tage_alt) .....	44
<b>5.5</b>	<b>Establishment variables</b> .....	<b>45</b>
5.5.1	Economic activity 93 generated – completed by extrapolation / imputation (w93_3_gen) .....	45
<b>5.6</b>	<b>Location data</b> .....	<b>45</b>
5.6.1	Place of work: region (ao_region).....	45
5.6.2	Commuter status (pendler).....	46
<b>5.7</b>	<b>Generated technical variables</b> .....	<b>47</b>
5.7.1	Source of the observation (quelle_gr) .....	47
5.7.2	Observation counter per person (spell).....	47
<b>6</b>	<b>References</b> .....	<b>48</b>
<b>7</b>	<b>APPENDIX</b> .....	<b>49</b>
<b>7.1</b>	<b>Overview of date adjustments and spell deletions</b> .....	<b>49</b>
<b>7.2</b>	<b>Comparison of the SIAB 7510 und SIAB-Regionalfile 7510</b> .....	<b>50</b>

<b>7.3</b>	<b>Recoding tables of aggregated variables in comparison with the weakly anonymous version of the SIAB 7510 .....</b>	<b>50</b>
<b>7.4</b>	<b>Frequencies.....</b>	<b>72</b>
<b>8</b>	<b>LIST OF ABBREVIATIONS.....</b>	<b>72</b>

## **Abstract**

This FDZ data report describes the “Regionalfile” of the Sample of Integrated Labour Market Biographies 1975-2010 (SIAB-R 7510). This dataset represents the factual anonymous version of the Sample of Integrated Labour Market Biographies (SIAB) and may be delivered to scientific research institutions after concluding a use agreement with the IAB.

## **Zusammenfassung**

Dieser FDZ Datenreport beschreibt das „Regionalfile“ der Stichprobe der Integrierten Arbeitsmarktbiografien 1975-2010 (SIAB-R 7510). Der Datensatz stellt die faktisch anonymisierte Version der Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) dar und wird nach Abschluss eines Nutzungsvertrages mit dem IAB an Forschungseinrichtungen übermittelt.

**Keywords:** German administrative micro data, labour market data, data manual, factual anonymized data, Scientific Use File

## **Disclaimer**

We would like to thank our colleagues in the Research Data Centre (FDZ) as well as the division for IT Services and Information Management (ITM) of the Institute for Employment Research for their cooperation and support. Our special thanks go to Konstantin Bätz, Ana Petrovic, Hannes Schild, Anja Wurdack and Marie-Christine Laible. This data report is a revision of the FDZ data report 07/2011 (authors: Dorner, König, Seth). Furthermore, parts of this data report are taken from the Data report on the weakly anonymous version of the Sample of Integrated Labour Market Biographies (SIAB) (FDZ data report 1/2013). Additionally, individual passages from IAB-internal data documentations by IAB ITM were adopted.

## Data availability

The dataset described in this document is available for use by professional researchers. Further information can be found on the website <http://fdz.iab.de>.

# 1 Introduction and outline

## 1.1 Introduction

The regional file of the Sample of Integrated Labour Market Biographies 1975-2010 (referred to as SIAB-R 7510 in the following) is a factually anonymous<sup>1</sup> scientific use file of the Sample of Integrated Labour Market Biographies 1975-2010 (SIAB 7510)<sup>2</sup>. The regional file is a specific type of scientific use file which distinguishes itself from other types of scientific use files by preserving a comparatively detailed regional data structure. However, it provides only a limited range of variables and values at the level of individuals and establishments for reasons of data confidentiality. Altogether the SIAB-Regional file in its current version covers the employment histories of 1.594.466 individuals. Their employment biographies are documented in a total of 41.390.318 lines of data.

The starting point for the production of the SIAB-Regional File is the weakly anonymous version of the SIAB 7510, which is a 2 % sample of the Integrated Employment Biographies (Integrierte Erwerbsbiografien - IEB) of the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung - IAB).

The IEB contains information which comes from different sources, and the individual's statuses are described exact to the day. They comprise all individuals who showed one of the following statuses at least once during the observation period (data source and the period covered in square brackets):

- employment subject to social security [Employee History (BeH); recorded from 1975 onwards]
- marginal part-time employment [Employee History (BeH); recorded from 1999 onwards]
- receipt of benefits in accordance with Social Code Book III [Benefit Recipient History (LeH); recorded from 1975 onwards] or Social Code Book II [Unemployment Benefit II Recipient History (LHG/XLHG); recorded from 2005 onwards]
- registered with the Federal Employment Agency [Bundesagentur für Arbeit – BA] as a jobseeker [Jobseeker History (ASU/XASU); recorded from 2000 onwards]
- participation in an employment or training measure [Participation-in-Measures History (MTH); recorded from 2000 onwards]

This FDZ data report describes the preparation of the SIAB-Regional File and the variables of the scientific use file. The data report is structured as follows: besides an introduction,

---

<sup>1</sup> For further information on the concept of factual anonymisation see Chapter 3.4

<sup>2</sup> For the weakly anonymous version of the SIAB 7510 see vom Berge et al. (2013)

Chapter 1 contains an outline of the data, the volume structure and a list of variables. A description of the individual data sources can be found in Chapter 2. Data preparation and the anonymisation concept of the scientific use file are described in Chapter 3. Chapter 4 discusses the data quality and Chapter 5 concludes with a description of the individual variables contained in the data set.

## 1.2 Outline

Current data version	Sample of Integrated Labour Market Biographies - Regional File 1975-2010 (SIAB-R7510).
Type and volume of the data	Factually anonymous data (scientific use file), 2% random sample of individual accounts drawn from the Integrated Employment Biographies (IEB) of the IAB.
Outline of the contents	<p><b>Employee History (Beschäftigten-Historik - BeH):</b> Annual notifications and end-of-employment notifications submitted to the social security agencies for employees covered by social security and employees in marginal part-time employment.</p> <p><b>Benefit Recipient History (Leistungsempfänger-Historik - LeH):</b> Information on benefit receipt in accordance with Social Code Book III (SGB III) for recipients of unemployment benefits, unemployment assistance and maintenance allowance.</p> <p><b>Unemployment Benefit II Recipient Histories drawn from A2LL (Leistungs-Historik-Grundsicherung aus A2LL - LHG) and XSozial-BA-SGB II (XLHG):</b> Data on individuals in receipt of basic social security benefits in accordance with Social Code Book II (SGB II) (co-operations of employment agencies and municipalities – ARGE, separate units responsible for the implementation of SGB II and authorised municipalities (getrennte Trägerschaft und zugelassene kommunale Träger).</p> <p><b>Jobseeker History (Arbeitsuchenden-Historik - ASU) and Jobseeker History drawn from XSozial-BA-SGB II (XASU):</b> Information on job-search activity (applicant pool of the Federal Employment Agency) including information on job-search activity reported via the transmission standard XSozial-BA-SGBII to the BA by authorised municipalities or separated responsibilities.</p>
Variables with regional relevance	333 districts and aggregated regions based on the place of work, categorical variable for commuter status/place of residence.
Date of territorial allocation	31.12.2010
Type of territorial allocation	All sources contain corrected territorial allocation.

Period covered	The period covered depends on the data source. BeH 1.1.1975 - 31.12.2010 ( 2008, 2009: 18-month file; 2010: 6-month file) LeH 1.1.1975 - 31.12.2011 (X)ASU 1.1.2000 - 31.12.2004 and 1.1.2007 - 31.12.2010 (X)LHG 1.1.2007 - 31.12.2010
Update frequency	Continuous
Data unit	Employees covered by social security (including marginal part-time employees from 1999 onwards), benefit recipients, jobseekers
Number of cases	1.594.466 individuals 41.390.318 observations without overlaps (after splitting)
Adjustments - outline	The data of the SIAB-Regional File went through several preparation and anonymisation procedures such as the recoding of individual variables, aggregations and deletions of variables and time periods after they have been drawn from the IEB.
Special features - outline	The data in the SIAB-Regional File went through an episode splitting procedure. For this, partly overlapping observations within an individual account are split in such a way that only non-overlapping or completely parallel periods are created.
File format and size	Stata (approx. 1.5 GB), SPSS (approx. 3.1 GB)
Data access	Data transmission solely to public research institutes or institutes which verifiably perform tasks defined as independent scientific research.
Degree of anonymisation	Factually anonymous data (scientific use file)
Citation method	Sample of Integrated Labour Market Biographies – Regional File 1975-2010 (SIAB-R 7510), Nuremberg 2013

Table 1: Overview

### 1.3 List of variables

The overview of variables lists the variable names and the longer descriptions of the variables. It also shows which variables are available for which data sources:

	Variable is available for the data source. Degree of completeness always > 0.85
	Variable is available for the data source. Lower or varying degree of completeness, see description of variable and frequency count

Variable is not available for this data source. Degree of completeness always < 0.05
---

Example: the school-leaving qualification (schbild) is only available for (X)ASU observations; the observations of the other data sources contain missing ".n" in this variable. Another characteristic of the SIAB is that some variables have different contents depending on the data source. For instance, for BeH observations the employment status variable (erwstat\_gr) contains the person group of the employment notification procedure, for LeH observations it contains the type of benefit, for (X)LHG observations it contains the SGB II status and for (X)ASU observations the job-search status. These differences are not immediately obvious from the variable name for every variable.

Type of variables	Variables	Page	1	2	3	4
			BeH	LeH	(X)ASU	(X)LHG
			197 5- 201 0	197 5- 201 0	1997- 2004 2007- 2010	2007- 2010
Identifiers	persnr	33				
	bnn	34				
Period of validity/ Dates	begorig	34				
	endorig	35				
	begepi	35				
	endepi	35				
Information on employment, benefit receipt and job-search	quelle_gr	47				
	erwstat_gr	43				
	grund_gr	<b>Fehler! Textmarke nicht definiert.</b>				
	tentgelt_gr	40				
Personal information	deutsch	<b>Fehler! Textmarke nicht definiert.</b>				
	gebjahr	<b>Fehler! Textmarke nicht definiert.</b>				
	frau	<b>Fehler! Textmarke nicht definiert.</b>				
	bild	36				
	beruf_gr	41				
	stib	42				
	schbild	38				
	tage_jung	44				
	tage_alt	44				
(Generated)	W93_gen_	45				

Establishment variables	gr				
Regional identifiers	ao_region	45			
	pendler	46			
Generated technical variables	spell	47			

Table 1: List of variables and degree of completeness

## 1.4 Volume Structure

Number of cases	before splitting	after splitting
BeH	25.533.525	28.532.729
LeH	3.693.753	5.159.497
X(LHG)	557.807	1.520.461
X(ASU)	3.746.678	6.177.631
Total number of observations	33.531.763	41.390.318
Individuals		1.594.466

Table 3: Volume structure

## 1.5 Changes as compared to SIAB Regional 7508

In comparison to the SIAB Regional 7508 some changes have been made. These changes are listed below.

### *Corrected territorial allocation*

All locations stated in SIAB 7510 incorporate the territorial allocation as of 31 December 2010.

### *Characteristics*

The level1 and level2 variables were no longer included in the variables of SIAB 7510 for data economy reasons. The variables can be generated by using the following (Stata-) commands, if needed:

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

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

During the observation period, the economic sector classifications changed multiple times, so analyses over a longer period of time are more difficult. The FDZ therefore developed methods to create time-consistent characteristics from the different economic sector classifications. A detailed description can be found in Eberle et al. (2011). The economic sector aggregates, as listed in SIAB Regional 7510, are based on the consistent classification WZ93.

## Anonymisation

It has been attempted to keep the anonymization procedures as close as possible to that of the previous version. However, some differences occur between the two versions, as aggregations or deletions were necessary. Thus, some variables may no longer contain certain values compared to the SIAB Regional 7508.

### 1.6 Data use

Upon request, the SIAB-Regional File is made available to researchers for a limited period of time for a project with defined contents, and applications may be processed on the premises of their research institution (§ 282 para. 7 SGB III).

In order to be able to use the data a project application must first be submitted to the FDZ. After approval by the FDZ a project specific data use agreement is concluded with the researcher's institution.

Details about different options of data access, our requirements for data use and application procedures are available on the FDZ homepage at <http://fdz.iab.de/en.aspx>.

## 2 Data sources

The administrative personal data was drawn from the Integrated Employment Biographies (IEB) of the IAB. The IEB unites data from five different data sources, each of which may contain information from different EDP procedures.

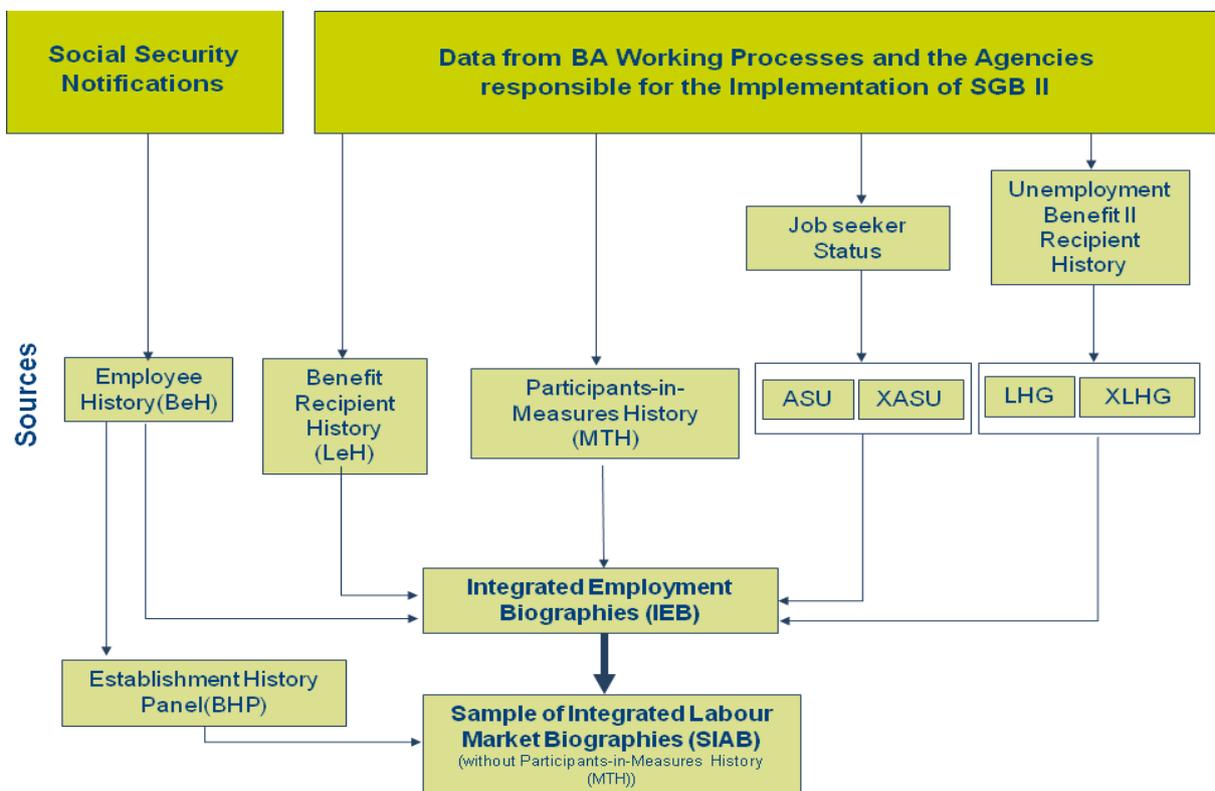


Figure 1: Data sources of the Integrated Employment Biographies (IEB)

The SIAB was drawn from the Integrated Employment Biographies (IEB) of the IAB as a 2% sample. Observations from all of the data sources shown, with the exception of participation (or planned participation) in employment and training measures (MTH), are incorporated into the SIAB from the IEB. Each data source of the IEB is specified below.

## 2.1 Employee History (BeH)

The data source regarding employment is the Employment History (Beschäftigten-Historik - BeH) of the IAB. The legal framework for data collection is the integrated notification procedure for health, pension and unemployment insurance, which came into effect as of 1 January 1973 (and was extended to cover East Germany as of 1 January 1991) and is known by the abbreviation DEÜV (previously DEVO/DÜVO) (for further details see: Bender et al. 1996, p. 4 ff.; Wermter/Cramer 1988). Under this procedure employers are required to submit notifications to the responsible social security agencies concerning all of their employees covered by social security. The BeH covers all white- and blue-collar workers as well as apprentices as long as they are not exempt from social security contributions. This means that civil servants, self-employed and regular students (see Cramer 1985) are in principle not recorded in the BeH. Since the notification procedure was changed on 1 January 1999, employees in marginal part-time employment and unpaid family workers have also been recorded (not contained in the data until 1 April 1999).

The data are recorded by the health insurance companies, collected in a continuous file by the BA and subsequently integrated into the History File of the IAB.

The variables of the SIAB-Regional file related to establishments are taken from the Establishment-History-Panel (Betriebs-Historik-Panel - BHP) which itself is also based on the BeH (see Hethy-Maier/Seth 2010). The following modified variables based on the BHP are included in the scientific use file:

- Place of Work (ao\_region)
- Industry Classification WS 73 (extrapolated until 2008) (w73\_gen\_gr)

Two issues have to be taken into account when linking individual data with establishment data:

1. Information on establishments is based on individual data and aggregated on the annual cut-off date of the BHP, June 30. Unlike the data on individuals, the establishment variables are not spell data but are only valid on 30 June precisely each year<sup>3</sup>.
2. Secondly, the BHP is created directly from the BeH, whereas the SIAB Individual File is generated from the IEB (see Figure 1). The IEB however does not incorporate all BeH observations (see Section 3.1).

## **2.2 Benefit Recipient History (LeH)**

The Benefit Recipient History (Leistungsempfänger-Historik - LeH) of the IAB covers periods during which individuals receive earnings replacement benefits from the Federal Employment Agency. The benefits comprise unemployment benefit, unemployment assistance and maintenance allowance, in other words not benefits under the jurisdiction of Social Code Book II (e.g. unemployment benefit II). Since the entitlement to receive benefits depends on meeting certain legal requirements, unemployment periods in which the requirements are not met (e.g. no entitlement for receipt of benefits in case of unemployment assistance, or non-completion of the qualifying period for unemployment benefit) are not reported in the Benefit Recipient History.

## **2.3 Unemployment Benefit II Recipient History (LHG / XLHG)**

The Unemployment Benefit II Recipient Histories (Leistungshistoriken Grundsicherung LHG / XLHG) contain the receipt of benefits in accordance with Social Code Book II (SGB II). This covers both basic social security benefits (e.g. unemployment benefit II) and supplements to unemployment benefit or additional benefits.

Unlike the benefits under the jurisdiction of Social Code Book III, the BA is no longer necessarily the sole institution responsible for administering the benefits. The data therefore distinguish between the three possible types of institutions responsible for implementing SGB II:

- Joint facilities (Gemeinsame Einrichtungen) / Cooperation of employment agencies and municipalities (Arbeitsgemeinschaften – ARGE, until the end of 2010) in which the BA and the municipality deal with tasks jointly,
- Separate responsibilities (getrennte Trägerschaft) (until the end of 2011) – the tasks continue to be divided between the BA and the municipality<sup>4</sup> –,

---

<sup>3</sup> An extreme example: there is an employment notification from 1.1.2006 until 30.5.2006; the establishment goes bankrupt in June 2006. There is then no information about this establishment in the BHP for 2006.

- Authorised municipalities which are also called opting local authorities or opting municipalities according to the initial experimental clause of Section 6a - here the local authority is responsible for all tasks in the sphere of SGB II.

The data of the “Unemployment Benefit II Recipient History drawn from A2LL” (LHG) come from the BA administrative procedure A2LL, whilst the “Unemployment Benefit II Recipient History drawn from XLHG-BA-SGB II” (XLHG) records data which was reported in accordance with the transmission standard X-Sozial-BA-SBG II.

As a rule, A2LL is used until 2010 in all ARGE cooperation projects, and from 2011 onwards in joint facilities, whilst the standard X-Sozial-BA-SBG II is used by the authorised municipalities. Both of the procedures are used by municipalities with separated responsibilities.

An important difference compared with the LeH is that the amount of benefits received is not determined at the level of the individual but at the level of the benefit community (Bedarfsgemeinschaft<sup>5</sup>). It is not possible, however, to link individuals with benefit receipt under SGB II at the level of benefit communities in the SIAB. The LHG and the XLHG in principle contain all registered individuals who are entitled to receive benefits in accordance with Section 7 SGB II, in other words, needy individuals who are capable of work and, if applicable, also the members of their benefit community. However, only periods during which a person is capable of work or is over the age of 65 are incorporated.

Each dataset depicts non-overlapping periods of entitlement to benefit of a person in a certain benefit community. A new observation starts for the following administrative reasons:

- on relevant birthdays (14, 15, 18 and 65) of the members of a benefit community, which are statutory and relevant for structural changes of the benefit community (see Chapter 3.1.5),

---

<sup>4</sup> The municipality pays the costs for housing and heating (Section 22 SGB II) and additional one-time benefit payments to cover extra costs (Section 23 (3) SGB II) as well as the additional benefits to support integration in accordance with Section 16 (2) Clause 2 No. 1 - 4 SGB II. The BA, on the other hand, covers the costs for regular benefits, social insurance contributions and integration benefits (SGB III and SGB II) and specific benefits excluding the additional benefits to support integration cited above.

<sup>5</sup> A so-called ‘Bedarfsgemeinschaft’, or ‘benefit community’, includes all individuals in a household who receive benefits jointly (i.e. as a joint payment). In the majority of cases, the ‘benefit community’ and the household will be the same, which in particular applies in the case of (married or unmarried) couples and parents with children under the age of 25. However, under specific circumstances the ‘benefit community’ might not include all household members, or a household in which everybody receives benefit payments might be made up of more than one ‘benefit community’. An example of the former is if a grown-up child lives with his/her parents and earns just enough to make his/her own living but has insufficient means to support his/her mother and father – in this case the ‘benefit community’ will only include the parents. An example of the latter is a three- (or more) generation household: since a ‘benefit community’ may only consist of two generations, this type of household will be made up of two such ‘benefit communities’, one consisting of the grandparent(s) and one of the parent(s) and child(ren) (Trappmann et al. 2007).

- if there are changes in the composition of a benefit community (e.g. by members joining or leaving the benefit community),
- if there are changes in the variables of the benefit community client, and
- at the beginning and end of a sanction period for observations from 01 April 2006 onwards. However it must be taken into account that it is not possible to determine either the duration, the type of the sanction or the time when the sanction was imposed or when it began based on the data. The reason for this is the lack of a corresponding variable or value that indicates the start, the type or the duration of a sanction.

The LHG data are available from 01 January 2005, the XLHG data from 01 November 2005. However, until the beginning of 2007 both data sources are incomplete (see Chapter 4).

## **2.4 Jobseeker Histories (ASU/XASU)**

Data about jobseekers are stored in the Jobseeker History (Arbeitsuchendenhistorik – ASU / XASU). The ASU data source contains information on jobseekers who are registered with employment agencies, and from 2005 onwards also includes ARGE cooperation projects and separated responsibilities for the implementation of SGB II. The XASU data source, on the other hand, contains the data of jobseekers in receipt of Unemployment Benefit II (ALG-II) from authorised municipalities from 2005 onwards. These data are reported in accordance with the X-Sozial-BA-SBG II standard.

# **3 Data preparation and sampling procedure**

## **3.1 Correction and validation procedures of the original data for SIAB-R 7510**

Before the data from the data sources specified in Chapter 2 are merged to form the IEB they undergo source-specific correction procedures (see the following chapters). The IEB as a whole undergo the following corrections:

- Observations in which the age is under 13 or over 75 are deleted.
- Observations whose end date precedes the start date are deleted.

Inconsistent information on gender or date of birth within an account is corrected; datasets without information on the date of birth are deleted. Further corrections, like in the IABS (completion procedure of presumably missing notifications, strike corrections) do not happen.

### **3.1.1 Employee History (BeH)**

- To capture a person group that is as constant as possible over time, some observations for which data are not available throughout the entire observation period are excluded. From 1999 onwards, these person groups are amongst others short-term employees,

freelance home workers, persons in workshops or in facilities of the youth welfare service, recipients of early retirement benefits, part-time farmers, compensation recipients, caregivers and notifications of the administration for military services and for alternative civilian services.

- Annual notifications of employees liable to social security (without special characteristics) with missing values in the income variable are not incorporated into the IEB.
- Gender and date of birth are taken from the social security number. This information is harmonised across data sources.
- The territorial allocations for place of work and place of residence are updated to the status as of 31 December 2010.

### **3.1.2 Benefit Recipient History (LeH)**

- In cases of on-going receipt the end date corresponds to the end of the period for which the benefit is allowed and not the end of the benefit receipt period. There is no information on the end of the period for which the benefit is allowed in the SIAB-Regional File because benefit receipt notifications still valid on 31.12.2010 are right-censored (see section 3.4.2). Observations with no end date or an invalid end date are excluded from further processing, since in these cases it cannot be assumed that a benefit payment was made at all.
- Observations without a valid start date are excluded.
- If the end date for unemployment assistance precedes the start date by one day, then the end date is corrected by one year.
- If the end date for the receipt of unemployment assistance precedes the start date by one day, then the end date is increased by one year.
- The territorial allocations are corrected analogous to the BeH.

### **3.1.3 Unemployment Benefit II Recipient History (LHG / XLHG)**

- It only contains observations of people who are capable of work and people over the age of 65.
- Individual-related variables that are available for the (X)LHG sources are valid for the entire observation period.
- The territorial allocations are corrected analogous to the BeH.

### **3.1.4 Jobseeker History (ASU / XASU)**

- There is no consolidation of the ASU observations of individual persons. Therefore, overlaps between ASU observations might occur.
- Individual-related variables that are only available for the (X)ASU sources always refer to the beginning of the spell.

- A new ASU spell is generated as soon as a change of status (e.g. from job-seeking to unemployed) occurs. This also applies if the type of institution (employment agency, joint facility, authorised municipalities, separated responsibilities) changes. The ASU data basis only distinguishes between observations with the status “unemployed”, “job seeker” as well as since 2006 “seeking advice” and “without status”. In the IEB, however, an additional status “ill / not able to work” is available. The employment status “ill / not able to work” is assigned to IEB spells when in the ASU data basis
  - a preceding observation with the status “unemployed” exists which joins the next observation without a gap and has “incapacitated for work” as the reason for exit, and
  - a subsequent observation with the status “unemployed” exists which also follows without a gap, and
  - the observation itself does not have the status “unemployed” but “jobseeker”.
- In contrast to the ASU source, the XASU only distinguishes the status
  - “not unemployed, but seeking job” or
  - “unemployed and simultaneously seeking job”.
- Unlike the ASU, periods of illness are not taken into account when generating the “employment status”, since there is no information about illness available in the XASU data. When calculating the unemployment duration with XASU observations, gaps due to illness cannot be identified.
- The XASU contains non-overlapping time periods for individuals. If one of the following variables changes, in each case a new data spell is generated for the XASU:
  - Change of job search status
  - Change of availability
  - Change of SGB II institution (due to notification procedure)
  - Change of place of residence
- The territorial allocations are corrected analogous to the BeH.

### **3.1.5 SGB II anonymisation**

In order to reduce the risk of de-anonymisation, in general only the year of birth is indicated in the administrative data. However, in the (X)LHG and (X)ASU sources there is the risk that the exact date of birth may still be obvious due to the chronological structure of the observations. Observations might end systematically on certain birthdays and/or the day before, or start again on the birthday.

In order to prevent an indirect determination of the exact day of birth, the following procedure is applied. Observations divided on the 18<sup>th</sup> or 65<sup>th</sup> birthday will be merged into a single one.

Besides the exact day of birth no other information will get lost in this case. For observations beginning on the 15<sup>th</sup> or 65<sup>th</sup> birthday (end on the 65<sup>th</sup> birthday), the start date (end date) of the observation will be set to the beginning (the end) of the respective quarter. This correction results in a bias of the duration of the observation.

Corrected observations are marked in the "employment status" variable.

### **3.2 Episode-Splitting**

The administrative individual data are available with "split" episodes. If observations overlap within an account, these observations are replaced by artificial observations with new dates so that completely parallel periods and non-overlapping periods are created. This increases the number of observations.

The original date variables for the beginning and the end of the original observation (begorig and endorig) are retained, the variables 'start date of the split episode' and 'end date of the split episode' (begepi and endepi) mark the beginning and the end of the split episodes. It is possible to establish whether observations have been split by comparing the original period (begorig and endorig) with the episode period (begepi and endepi).

To restore the original data without the split episodes or to delete the episodes that were created artificially by means of episode splitting it is necessary to select all observations for which the start of the original observation is the same as the start of the split episode (begepi == begorig).

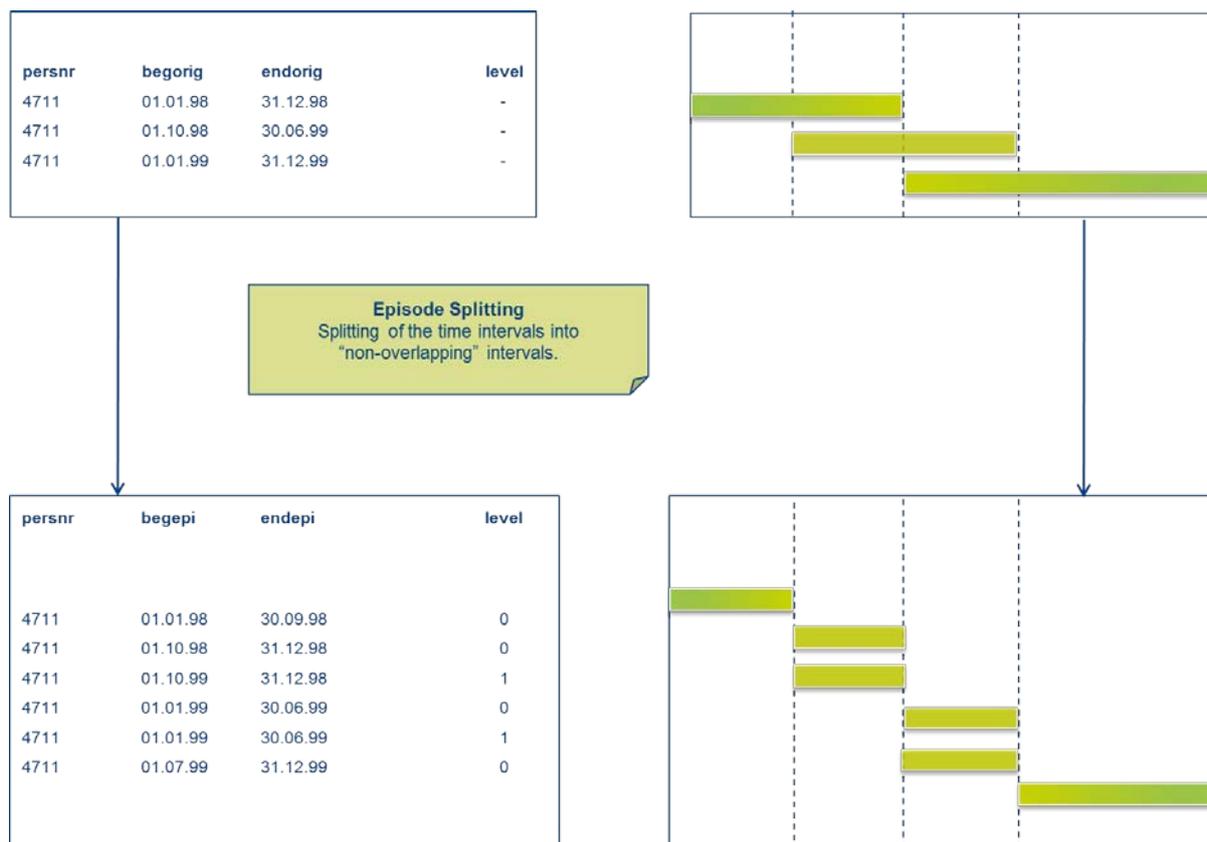


Figure 2: Episode-splitting

### 3.3 Sampling procedure

The SIAB is a 2% sample of the individuals for whom there is an entry in the IEB. This makes the SIAB representative of the person groups of all of the data sources, not only of people in employment. Accordingly, case numbers of the population can be estimated on the basis of the sample with a weighting factor of 50. This applies in principle to the person groups of all of the data sources; however, the differing degrees of completeness of the data sources have to be taken into account.

### 3.4 Anonymisation of the SIAB-R 7510

The Sample of Integrated Labour Market Biographies, which serves as the data basis for the SIAB-Regional File, contains a large amount of personal information on the status of a person on the labour market. Additional establishment characteristics are available in the case of employment notifications from the social security notification procedure. The legislator stipulates that these social data (§ 67 SGB X) which are available at micro level at the FDZ may only be transmitted for scientific purposes (§ 282 para. 7 SGB III, § 75 SGB X) if both the personal and the establishment-related data can be regarded as factually anonymous when tested. This means that a disproportionately large amount of time, expense and effort

would be required in order to identify an individual in the data set. The costs of deanonymisation must therefore exceed its benefits.

The guidelines provided by Müller et al. (1991) regarding the factual anonymity of micro data constitute the decisive orientation framework for the preparation of scientific use files at the FDZ (see for example Zimmermann et al. 2007; Drews 2008). The Sample of Integrated Labour Market Biographies was prepared and anonymised in the form of a regional file in order to guarantee the best possible continuity for users of previous scientific use files, in particular for users of the IAB Employment Samples, and in order to meet the constant demand for a regional file of employment history data from the IAB. The data were anonymised by deleting selected values and variables and aggregating information which is sensitive with regard to data protection legislation. The anonymisation procedures are described in the following.

### **3.4.1 Merging sources**

When developing the SIAB-Regional File from the SIAB 7508, details on the source of certain observations were merged. For example, information from the original data sources of the Unemployment Benefit II Recipient Histories (LHG and XLHG) and information about job-search (ASU and XASU) were combined in each case to a joint data source, (X)ASU and (X)LHG, in the SIAB-Regional File. The reason for this is that the information about whether a person in a region receives his/her support from an authorised municipality (zugelassener kommunaler Träger - zKT) constitutes a risk of the data subject being re-identified. This risk results from the fact, that the information about the institution responsible for the person, in combination with regional information from other sources or from the employment history can make it possible to identify the location at local authority level.

### **3.4.2 Periods of validity of the sources**

After aggregating the data sources (see Section 3.4.1) and in accordance with the data protection regulations mentioned above, information about the start and end dates of selected observations were adjusted. Figure 3 illustrates the periods for which the various sources are available in the SIAB-Regional File.

For the Employee History (BeH) and the Benefit Recipient History (LeH) the SIAB-R 7508 only contains observations for the period between 1.1.1975 and 31.12.2008. Notifications submitted in the context of the social security notification procedure in East Germany are only assumed to be complete from 1992 onwards. In employment notifications with valid places of work in East Germany before the year 1992 the information on the district region is set to missing (.z). It is therefore only possible to distinguish whether a person is employed in East or West Germany in the SIAB-Regional File from 1992 onwards.

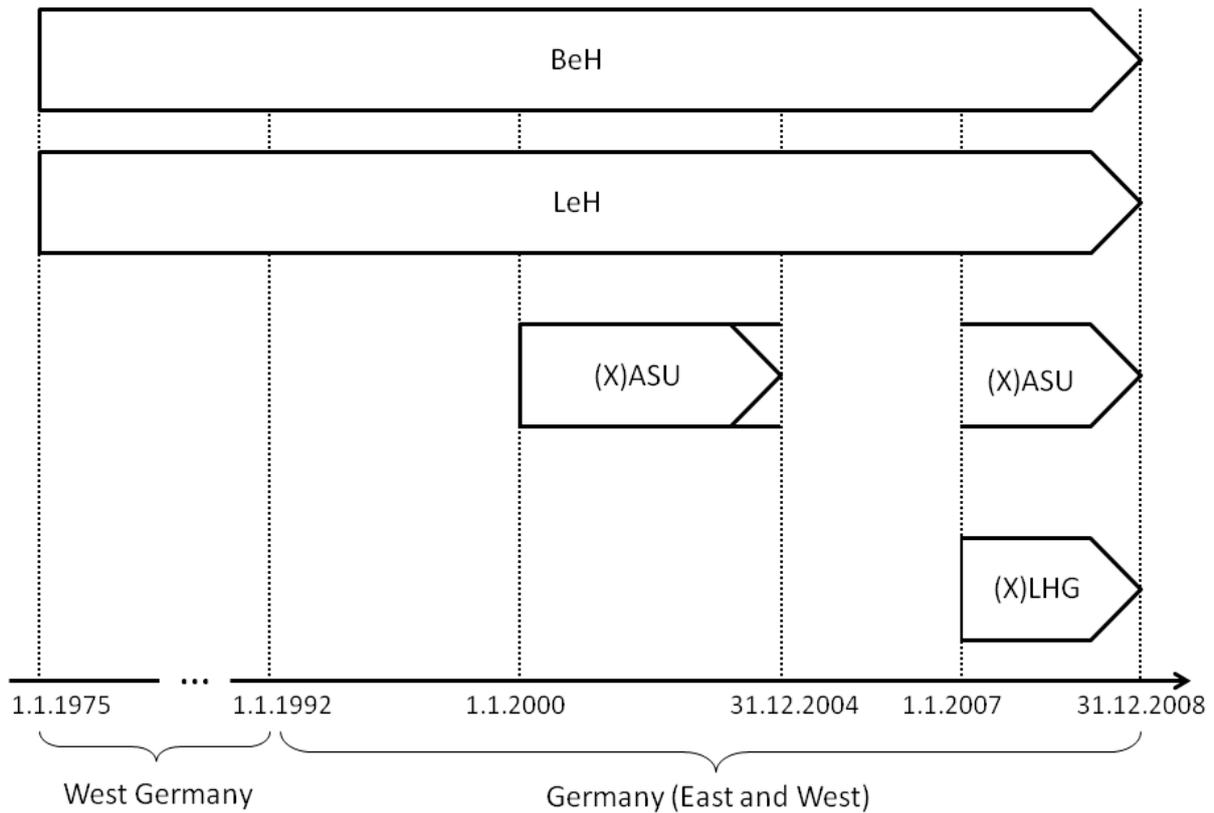


Figure 3: Periods of validity of the data sources

The (X)ASU is available in the SIAB-Regional File for the period between 1.1.1997 and 31.12.2004 and the period between 1.1.2007 and 31.12.2010. As the ASU original source is only assumed to be complete from the year 2000 onwards, observations with an end date before 1.1.1997 are not reported in the SIAB-Regional File. The deletion of observations in the years 2005 and 2006 is first due to the quality-relevant under-representation of notifications from the authorised municipalities responsible for implementing SGB II (zugelassene kommunale Träger) in the IEB original data (see vom Berge et al. 2013). Second, revealing these data is problematic in terms of data protection legislation because of the risk of individuals reported by authorised municipalities being identified in combination with regional information. The outlined problems associated with notifications from authorised municipalities also apply to the (X)LHG in 2005 and 2006. Accordingly the SIAB-Regional File only contains observations from the (X)LHG with episode start dates from 1.1.2007 onwards.

Observations whose start and end dates are outside these source-specific periods of validity are no longer contained in the data. In the case of observations from the LeH, (X)ASU and (X)LHG sources which span more than one year and in which either the start date or the end date falls in one of the source-specific periods of validity, the dates are changed. This means

that the original information regarding either the start or end date is replaced by the respective start or end date of the source-specific period of validity.

(X)ASU records constitute a peculiarity in the data. They can begin before the deleted period of 2005/2006 and can extend into 2007 without interruption. These observations were duplicated in the SIAB-Regional File and the start and end dates were adapted to the period of validity, so the data contain two observations – a right-censored record and a left-censored one.

The following fictitious examples in Figure 4 illustrate the procedure used for adjusting the dates. The white bar [1] always represents the original period and the grey bar [2] the period contained in the SIAB-Regional File after the date adjustment procedure.

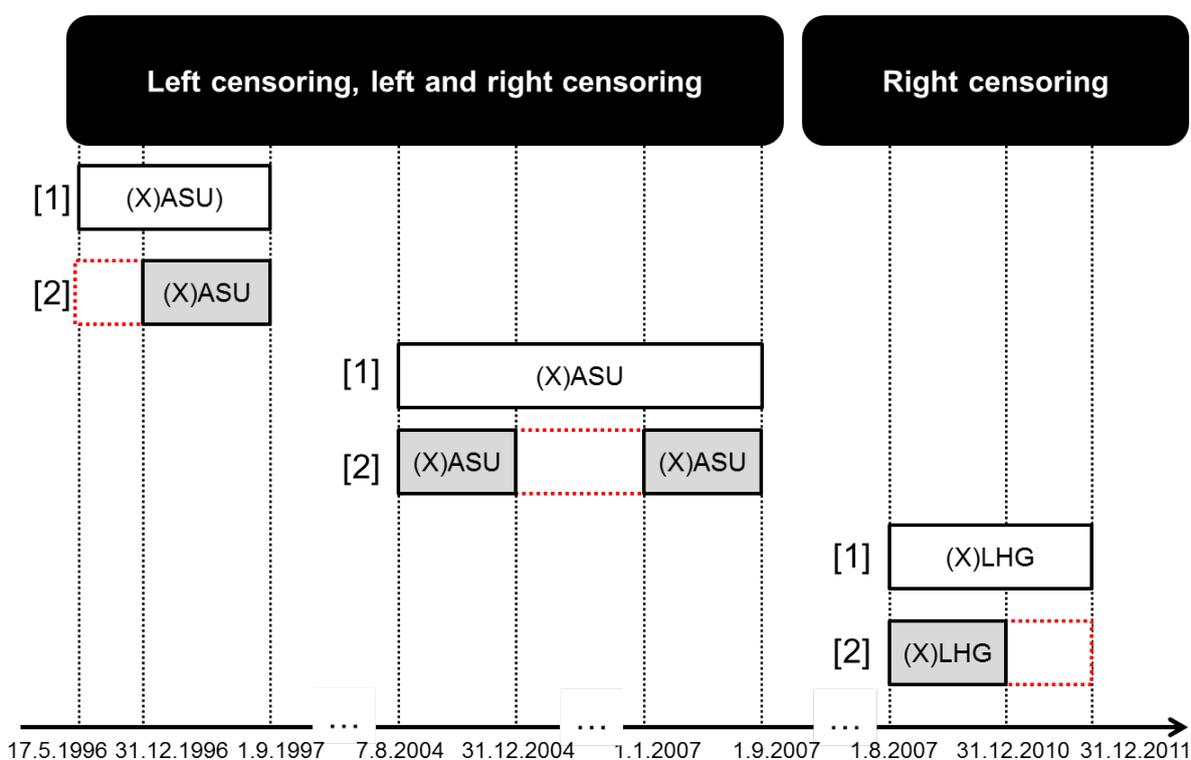


Figure 4: Examples of the date adjustment procedure for observations spanning more than one year

In the example showing the left-censoring of a job-search observation from the (X)ASU to 1.1.1997 and in the example depicting the right-censoring of an observation from the (X)LHG (also LeH or (X)ASU) to 31.12.2010, the number of data records in the data remains the same. Only the parts of the multi-year observation which are outside the observation period covered by the SIAB-Regional File are deleted.

The example in the middle shows the procedure used for multi-year observations which span the period deleted from the (X)ASU in 2005 and 2006. In this case two censored observa-

tions are created from the original notification. Accordingly the number of lines of data in the SIAB-Regional File increases compared with the original file.

An overview of the date conversions carried out can be found in tabular form in the Appendix (see Section 6, Table A2).

### 3.4.3 Censoring of data on age

For data confidentiality reasons the scientific use file only contains those data records in which a person's age is between 17 and 62. Records that end before the age of 17 or begin after the age of 62 are excluded. In the case of records spanning more than one year which begin before or in the year of a person's 17th birthday but end after that date, the original start date ('begorig') was converted to 1<sup>st</sup> of Januar of the year in which the person reaches the age limit of 17. Similarly, the original end date ('endorig') was converted to 31 December of the year in which the person reaches the age limit of 62 if the record begins before or in the year of a person's 62nd birthday and end after that year. The variables 'tage\_jung' and 'tage\_alt' report the number of days in (regular) employment<sup>6</sup> before the age of 17 and after the age of 62, respectively.

### 3.4.4 Aggregation and deletion of variables

While in the case of the sources with information on benefit receipt and job-search activity the spectrum of variables was strongly reduced compared with the weakly anonymous version of the SIAB for data protection reasons, the employment-related variables from the Employee History (BeH) have largely been retained and aggregated in the SIAB-Regional File.

The variables remaining in the SIAB-Regional File constitute the basis for the application of further anonymisation procedures in the form of the aggregation of information. Müller et al. (1991, p. 444 f.) suggest the following three rules for testing and creating a factually anonymous regional file:

1. *It must not be possible to identify any regional unit contained in the data which has fewer than 100,000 inhabitants by combining regional classifications (see Müller et al. 1991, p. 444)*

This demand was met by checking the number of inhabitants in the 413 autonomous municipal districts (kreisfreie Städte) and rural districts (Landkreise) as of 31.12.2010<sup>7</sup>. The region-

---

<sup>6</sup> The calculation of periods of employment takes the following restrictions into account: `quelle_gr==1 & level2==0` (Stata Syntax).

<sup>7</sup> Data basis: Population projection of the Federal Statistical Office (Statistisches Bundesamt) as of 31.12.2010 for the autonomous municipal authorities (kreisfreie Städte) and the administrative districts (Kreise) (NUTS 3-level) of the Federal Republic of Germany.

alisation contained in the SIAB-Regional File combines the information on districts in the original data in such a way that each of the district regions created has no fewer than 100,000 inhabitants as of the reference date. This regionalisation in the SIAB-Regional File is based on the procedure used to create the microcensus district regions (Mikrozensuskreisregionen - MZKR) (see Christians/Wirth 2009, p.22). The aggregation results in a total of 332 district regions which are shown in the variable 'place of work: region' (ao\_region) (territorial allocation 31.12.2010). When defining the district regions superordinate spatial units such as spatial planning regions according to the Federal Institute for Research on Building, Urban Affairs and Spatial Development (Bundesinstitut für Bau-, Stadt- und Raumforschung - BBSR) and NUTS-II regions were taken into account as far as possible (see Section 7.3, Figure A1/Table A8).

In addition the variable 'commuter status' (pendler) was generated, which indicates whether the current place of residence differs from the current or most recently reported place of work.

2. *Variables that may potentially be suitable for identifying individuals must be so strongly aggregated that no values of variables are shown which*
  - a. *do not comprise at least 50,000 individuals in the population,*
  - b. *do not contain at least three individuals in the data set per regional unit transmitted; variable values which contain only one or two cases in the data set may only be shown in a more strongly aggregated form (see Müller et al. 1991, p. 444 f.).*
3. *If necessary for compliance with data protection legislation, all other variables contained in the regional file must be aggregated to a level at which every variable value disclosed covers at least 5000 persons in the population (see Müller et al. 1991, p. 445).*

On the basis of these criteria, further variables were aggregated when producing the SIAB-Regional File. Frequency counts of individuals as of 30th June of any year were taken as the basis. Attributes of variables with a high risk of deanonymisation were aggregated first if their univariate marginal distribution as of 30th June of the respective year did not comprise a minimum number of persons (2a). Second, they were aggregated if the cells of the contingency tables showing the relationship between the particular variable and the variable 'place of work: region' did not contain a minimum of three persons (2b). When checking the data on economic activity, the number of establishments was taken as the basis and not the number of persons as this is an establishment variable. All other variables are counted over the entire observation period of the data and are aggregated if necessary on the basis of (3) if at any

time the minimum case numbers given above are not reached in the data cells of the SIAB 7510.

If parallel information from different sources is available for a person on the reference date, e.g. job-search and simultaneous benefit receipt, then the person was only taken into account once in a variable that is available in both sources (e.g. age or nationality) when conducting the frequency counts covering more than one data source.

Aggregations were conducted for the following variables: nationality, occupation – activity performed, classification of economic activity, occupational status, reason for notification/ reason for end of benefit receipt/ reason for discontinuation of unemployment benefit II/ reason for deregistration. For these variables the years 1975 up to and including 1984 were not used to assess the aggregation level required for the factual anonymity of the variables. Due to the structural change on the labour market, a strict interpretation of the anonymisation rules described above would have resulted in a very high level of aggregation in this period and therefore in a substantial restriction of the analysis potential for the data for certain groups of persons. The daily wage and the benefit rates were rounded to whole Euros.

### **3.4.5 Consequences of the anonymisation**

The information on employment and benefit receipt in accordance with SGB III in the SIAB-Regional File are only affected to a small extent by deletions at the level of data records and individuals. The deletion of individual data records for anonymisation reasons in the process of adjusting the observation periods for the (X)ASU and (X)LHG sources resulted in some individuals being dropped from the SIAB-Regional File. As a result of the data preparation procedures mentioned above a total of 44.859 individual accounts were deleted with reference to the SIAB 7510. The SIAB-Regional File therefore still contains 97.3 percent of the persons included in the source file, SIAB 7510 (see Section 7.2, Table A2).

Some personal and establishment-related variables were either aggregated or deleted from the data during the anonymisation process across all sources. In the data sources with information on benefit receipt and job-search, the spectrum of variables was reduced considerably, while the employment-related variables were largely retained and were simply aggregated. The SIAB-Regional File thus provides the continuity for employment-related analyses to a large extent compared with the IAB Employment Samples (especially the IABS-Regional File 1975-2004).

More detailed information about deletions and aggregations at variable level due to anonymisation compared to the weakly anonymous version of the SIAB 7508 and compared to related FDZ data products can be found in the tables in the Appendix to the data report.

### 3.5 Missing values

In the SIAB-R 7510 missing values or values that are invalid due to anonymisation are coded as follows:

Term	Value	Description
No (valid) details available, anonymised values	.z	Values of a variable which are not systematically missing, i.e. the variable is available in principle for the data source, but no details are available for the value considered or no allocation is possible.  Otherwise: anonymised values due to data confidentiality obligations.  Please note that it is not possible to distinguish between anonymised and invalid values originating from the weakly anonymous version of the SIAB.
Systematically not available	.n	A variable is not available in principle for a data source (dark grey cells in the overview of variables in Section 1.3) or is not available for a certain period.

## 4 Data quality and problems

The IEB contains employment histories. Not each kind of employment is included in the administrative data, though. Certain persons with certain life courses are not represented in the IEB at all.

For evaluation purposes, it is often relevant to know gaps in the included biographies (e.g. for control group creation, course of life analyses, etc.) The gaps listed in the following are defined as periods of time after the end of school education for which no data is included in the IEB. These gaps can be divided into

- gaps with no information available, and
- gaps for which information may be available from the reason for notification / reason for end of benefit receipts / reason for discontinuation of SGB II / reason for deregistration characteristic of the direct preceding dataset (if a corresponding dataset exists).

These gaps were identified using the manifestation of the reason for notification / reason for end of benefit receipts / reason for discontinuation of SGB II / reason for deregistration and employment status characteristics in the various sources of SIAB 7510. The list makes no claims of being complete. The anonymization of this variable in the SIAB Regional File further reduces identifiability.

<b>Biographical gap</b>	<b>Information on gap, identifiable using the manifestation of the "grund" characteristic in the preceding dataset of the source, if necessary</b>
Civil servants, professional soldiers, judges, employees of bodies or foundations under public law, unsupported self-employed persons	LeH, ASU
Students, people in school-based further education	LeH, ASU
People who are ill / not able to work for more than 6 weeks (illness during unemployment, however, is represented in the ASU source under certain conditions, see Chapter 3.1.4)	BeH, LeH, ASU
People receiving old-age pension without employment if not member of a benefit community	LeH, ASU
Recipients of early retirement benefits	LeH, ASU
Trade professionals working from home	
Short-time employed persons	ASU
People in facilities of the youth welfare service organisation, in vocational training centres, approved workshops or similar facilities for disabled persons	ASU
Participants in benefit programmes for participation in working life (people in rehabilitation)	ASU
(Sideline) farmers	
Caregivers according to Section 19 SGB XI	
Conscriptors	BeH, LeH, ASU
Persons in reserve duty training	BeH, LeH, ASU
Persons fulfilling community service	BeH, LeH, ASU
Persons fulfilling a voluntary social or ecological year instead of community service	
Other people not registered with the statutory pension insurance or the Federal Employment Agency (e.g. sabbatical, funding from personal assets or pensions, emigration, employment abroad, voluntary work, etc.)	LeH, ASU, BeH
Strikers in case of a duration of the strike of more than a month	LeH
Social benefit recipients (prior to the introduction of SGB II in 2005), social allowance recipients (according to SGB II)	
Compensation recipients according to FELEG (Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit, Act on the Support in Case of Termination of Farming Activities)	
People participating in measures	ASU

Table 4: Biographical gaps and identification possibilities

## 4.1 Employee History (BeH)

- Due to the introduction of the employment notification procedure in the federal states of Eastern Germany, the notifications for Eastern Germany can only be assumed to be sufficiently complete from 1993 onwards. For the same reason, a great deal of spells in 1991 show missing values for several variables (such as employment status).
- The increase in the number of BeH observations from 1999 onwards is due to the introduction of the obligation to submit employment notifications for people in marginal part-time employment from 01 April 1999 onwards.
- Especially in 1999, observations with part-time employment increased significantly. This is caused by the actually observed increase in part-time work as well as by the fact that since 1999 employment notifications have been completed more properly.
- Within the employment notification procedure, a certain time lag is unavoidable. Although changes in employment relationships have to be reported immediately, and existing employment relationships have to be confirmed annually by April of the following year, some notifications actually arrive years later. The History File of the IAB is not updated continuously, however, but at certain intervals. This is done using files of employment notifications for one particular year which were submitted 36, 18, 12 or 6 months after the end of the reporting year (e.g. the 18-month file for 2007 can be created in July 2009 at the earliest). Notifications submitted more than three years late are not taken into account at the IAB, which means that a 36-month file shows a 100 % degree of completeness by definition.
- In the version of the IEB on which the SIAB data are based the degree of completeness of the BeH observations last stood at 100 % in 2007. 18-month files were used for 2008 and 2009, and the observations for 2010 come from a 6-month file. It can therefore be assumed that the BeH observations for 2008, 2009 and in particular for 2010 are slightly underreported in the SIAB. However, this should not reduce the ability to analyse the data at the individual level. The missing notifications occur more frequently in a few establishments, however. This means that in individual cases the establishment data, e.g. establishment size, are grossly incorrect and will change considerably in subsequent versions.
- In 1984, a change was made in the employment notification procedure. From that time onwards, one-off payments of gross earned income were reported as part of the annual earnings subject to social security contributions, which lead to an increase in the average daily wage. In particular, the proportion of wages and salaries above the upper earnings limit has increased considerably from that year onwards (see Bender et al. 1996).
- For the years 1992 until 2000 noticeable drops and rises in the number of notifications were perceived. Drops can be observed especially for the following 10 districts: Braun-

schweig (03101), Wolfsburg (03103), Emden (03402), Kassel (06633), Essen (05113), Neuss (05162), Erftkreis (05362), Hersfeld-Rotenburg (06632), Miltenberg (09676) and Kempten (Allgäu) (09763). This is due to notification problems of a single or several establishments in these regions.

- Considerable drops were also observed for the districts Salzgitter (03102) und Hoyer-swerda (14264).
- Concerning the notifications for full-time employment, especially the districts Main-Taunus (06436) and Alzey-Worms (07331) are noteworthy. They feature above-average rises. Also in this case, the reasons are notification problems of a single or several establishments in these regions.
- In the years 1996 to 1998, the value 102 (doctors and pharmacies) within the "occupation – activity performed" variable are very rare compared to the neighbouring years.
- For single birth cohorts, the number of cases for 1975 is lower than expected. Therefore, it is problematic to analyze the data for the year 1975 (see Bender et al. 1996).
- Before 1999 additional notifications without (valid) information about the person group occur. This is particularly relevant for the year 1991, as many employees from East Germany are reported in this way in 1991.

#### **4.2 Benefit Recipient History (LeH)**

- For the states of Eastern Germany, the LeH observations were not fully recorded until 1992.
- The benefit receipt data used to be saved on magnetic tapes. Owing to a fault in one magnetic tape, the benefit receipt data up to and including 1980 are only partially available. Thus, in the present data product, it can be assumed that information on benefit receipt in that period is not available in full as well.
- Caused by an internal change of systems, there is a break in the collection of periods of exclusion of benefits and of benefit suspension in 2004 due to failure to comply in 2004.

#### **4.3 Unemployment Benefit II Recipient History (LHG/XLHG)**

- With regard to the completeness of case numbers or benefit histories from the LHG and XLHG data sources, there are substantial gaps in the years 2005 and 2006. These years are censored in the SIAB Regionale File (see Section 3.4.2).
- Also as of 2007, underrecording of mostly one month depending on the institution has occurred, especially in the XLHG. The LHG shows underrecording and overrecording.
- Some individuals for whom a (X)LHG spell exists are excluded completely or partly from the benefit receipt according to SGB II, for instance because they are in a subsidised training programme, receive old-age pensions, live in a stationary institution or receive in-

surance payments to avoid need for assistance. This affects 3 to 5 percent of all cases on average.

- The degree of completeness of the variable “Reason for discontinuation of unemployment benefit II” (grund\_gr, see Section 5.4.1) is comparatively low in the SIAB data. However, it must be taken into consideration here that the “Reason for discontinuation of unemployment benefit II” is a purely person-related variable. Thus if an observation of an individual in receipt of Unemployment Benefit II is split due to a change in the structure of his/her benefit community, the variable “Reason for discontinuation of unemployment benefit II” shows the value “missing” (.z).
- Although the beginning and the end of a benefit sanction lead to the observations being split, it is not possible to determine from the data either the duration or the type of the sanction or the time when the sanction was imposed or when it began.

#### **4.4 Jobseeker History (ASU/XASU)**

##### **4.4.1 ASU**

- The registered periods of job search activity in the ASU source have been regarded as complete from the year 1997 onwards. Therefore, the analysis potential of the ASU spells before 1997 is limited. For the Scientific Use File ASU data sets for the years before 1997 have been deleted.
- For some people for whom an opting municipality has been responsible since 2005, "artificial" ASU datasets were created by the Federal Employment Agency in parallel. These can be identified via estatvor (transfer to an authorised municipality).
- From mid-2005 until mid-2006, the coArb IT procedure, from which the jobseeker and applicant pool data originate, was superseded by the VerBIS procedure at the Federal Employment Agency. The information for many of the variables recorded was gathered with different levels of differentiation and different qualitative weighting in the two systems. It is therefore very difficult to integrate these variables into the IEB, which is only possible using a special procedure (mapping). Unfortunately, a full conversion of the affected variables from coArb to VerBIS cannot be achieved by means of mapping, so for some variables there is an unusually large number of the values “no details available”, “other” or “missing”. Moreover, striking differences may occur in frequency counts, depending on whether the original source of the data was coArb or VerBIS. Important limitations of the analysis potential are mentioned in the corresponding description of variables.

#### 4.4.2 XASU

- In the scientific use file the XASU notifications from the 69 authorised municipalities are assumed to be complete in the IEB as of the year 2007 onwards.
- As there is no central allocation of identification numbers, inconsistencies or implausible job-search histories are possible in the IEB and therefore also in the SIAB data.
- Several variables in the Scientific Use File sometimes have only a very low degree of completeness for the XASU. Variables which are affected by this include “nationality”, “reason for notification”, “employment status”. The degree of completeness of these variables improves over time. In general it is to be taken into account that the degrees of completeness sometimes vary depending on the institution supplying the data.
- It remains unclear whether the institutions – as intended – only report periods in which they themselves are responsible for the jobseekers or whether preceding periods of support from the BA in accordance with SGB III are reported as continuous episodes if the job-search status does not change.
- For a number of institutions the proportion of registered recipients of unemployment benefit II who are also registered jobseekers is implausibly large at times or continuously in the IEB. One possible reason for this could be an incorrect determination of the status “not unemployed but seeking work” by these institutions.
- The institution-related and period-related plausibility of the XASU data should be examined before use, taking the research question into account.

## 5 Description of variables

Frequency counts and overviews of the individual values and labels of the variables can be found in separate files under <http://fdz.iab.de>.

### 5.1 Identifiers

#### 5.1.1 Systemfreie Personnummer (persnr)

Variable label	artificial individual ID
Variable name	persnr
Category	identifiers
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	numerical
Hierarchy	none
Detailed description	Indicates which observations belong to the same person. Artificial means that it is not possible to infer any of the person’s characteristics or any original identifiers from this individual ID. As there is no uniform individual identifier in the different data sources,

	<p>the allocation of the information from different data sources (e.g. employment and benefits) to individuals is not always unambiguous. In such cases, implausible employment histories may arise.</p> <p>The formation of the individual identifier which spans all data sources is based on a heuristic developed by the BA.</p> <p>The artificial individual ID from the SIAB 7508 was replaced by an artificial, randomly assigned individual ID in the SIAB-Regional File.</p>
--	---

### 5.1.2 Establishment counter (bnn)

Variable label	artificial individual ID
Variable name	persnr
Category	identifiers
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	numerical
Hierarchy	none
Detailed description	<p>Numbers the establishments in a person's working life in ascending order.</p> <p>Example: The first establishment in which a person was employed is given the value 1. If the person moves to a different establishment, this establishment is given the value 2, etc. If the person returns to an establishment in which he or she was previously employed, then this establishment is given the value that applied for the first period of employment there (e.g. 2). If a person returns to the first establishment after just one change of establishment, this would result in the sequence 1-2-1 for the variable 'bnn' over time.</p> <p>Missing establishment numbers in the original data were also set to missing (.z) in the SUF.</p>

## 5.2 Period of validity

### 5.2.1 Original start date of observation (begorig)

Variable label	original start date of observation
Variable name	begorig
Category	period of validity
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	date
Hierarchy	none
Detailed description	<p>Corresponds to the original start date of the notification. This can differ from the start date of the episodes (begepi) (see also the comments on episode splitting in Chapter 3.2).</p> <p>Since the notification logic might under certain circumstances allow for re-identification of the exact day of birth, the original information on the date were changed in these cases by applying the anonymisation procedure described in Chapter 3.4</p> <p>Because of the rules of the notification procedure, in BeH observations the starting and ending year are always identical (obligation of the employer to submit annual employment notifications).</p>

## 5.2.2 Original end date of observation (endorig)

Variable label	original end date of observation
Variable name	endorig
Category	period of validity
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	date
Hierarchy	none
Detailed description	<p>Corresponds to the original end date of the notification. This can differ from the end date of the relevant line of data, the so-called end date of the split episode (see also the comments on episode splitting in Chapter 3.2).</p> <p>Since the notification logic might under certain circumstances allow for re-identification of the exact day of birth, the original information on the date were changed in these cases by applying the anonymisation procedure described in Chapter 3.4.</p> <p>Because of the rules of the notification procedure, in BeH observations the starting and ending year are always identical (obligation of the employer to submit annual employment notifications).</p>

## 5.2.3 Start date of split episode (begepi)

Variable label	start date of split episode
Variable name	begepi
Category	generated period of validity
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	date
Hierarchy	none
Detailed description	It always equals or is greater than the start date of the original observation (see also the comments on episode splitting in Chapter 3.2).

## 5.2.4 Start date of split episode (endepi)

Variable label	start date of split episode
Variable name	begepi
Category	generated period of validity
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	date
Hierarchy	none
Detailed description	It always equals or is greater than the start date of the original observation (see also the comments on episode splitting in Chapter 3.2).

## 5.3 Personal information

### 5.3.1 Gender (frau)

Variable label	gender
Variable name	frau

Category	personal variable
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	numerical
Hierarchy	none
Detailed description	Gender dummy (0 - man, 1 - women). The gender information is constant within one individual account.

### 5.3.2 Year of birth (gebjahr)

Variable label	year of birth
Variable name	gebjahr
Category	personal variables
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	numerical
Hierarchy	none
Detailed description	Constant within one individual account.
Notes on quality	In the original data, it may happen that the year of birth changes between the data sources. This is corrected during the data preparation process. The information from the social security number is given highest priority here.

### 5.3.3 Nationality (nation)

Variable label	nationality
Variable name	nation
Category	personal variables
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	numerical
Hierarchy	none
Detailed description	The variable contains the nation codes used by the Federal Statistical Office (Statistisches Bundesamt).
Notes	Reports whether the person is a German citizen (1) or whether he or she is not (0). The notification may change over time.

### 5.3.4 School education and vocational training (bild)

Variable label	school education and vocational training
Variable name	bild
Category	personal variables
Origin	BeH, ASU, XASU
Data type	numerical
Hierarchy	none

Detailed description	<p>The variable has a different meaning depending on the data source:</p> <p><b>1) BeH</b></p> <p>Contains the school education / vocational training reported by the employers in the employment notification procedure. Combined variable which is used to gather information on both the employee's school education and his/her vocational qualifications (the highest qualification gained in each case). Here the school qualifications are to be established first and then the vocational qualifications. Only in the case of the values 5 (degree from a university of applied sciences) and 6 (university degree) are no other vocational qualifications determined (see BA 2005, p. IX.). The following values exist:</p> <ol style="list-style-type: none"> <li>1 Primary school / lower secondary school / intermediate school leaving certificate or equivalent school education, without a vocational qualification</li> <li>2 Primary school / lower secondary school / intermediate school leaving certificate or equivalent school education with a vocational qualification</li> <li>3 With upper secondary school leaving certificate (Abitur), without a vocational qualification</li> <li>4 With upper secondary school leaving certificate (Abitur) and a vocational qualification</li> <li>5 Degree from a university of applied sciences (Fachhochschule)</li> <li>6 University degree</li> </ol> <p>It cannot be inferred from the data whether the individual completed school education or not. Both cases can occur in the values "1" and "2".</p> <p>"Changes in the vocational training status frequently occur at the same time as a change of establishment. This is because the notification data are compiled anew in the new firm. If, for example, an employee has gained a higher qualification via a part-time further training course while still working then this change of status is probably not recorded until he/she joins a new firm. It can generally be assumed that when a person is employed in a firm for a longer period, the personal data that they reported when they joined the firm is simply continued." (Meinken / Koch 2004, p. 63).</p> <p>A method for correcting missing values or temporal inconsistencies in the education and training data in the predecessor sample, the IABS, can be found in Fitzenberger et al. (2006) and in Drews (2006). Only the BeH data source is used for this, however, as this was the only data source with information on education and training in the IABS.</p> <p><b>2) ASU / XASU</b></p> <p>Contain the completed vocational training at the beginning of the job search activity. The following values exist:</p> <ol style="list-style-type: none"> <li>21 no completed vocational training</li> <li>22 in-firm training</li> <li>23 external training</li> <li>24 full-time vocational school (Berufsfachschule)</li> <li>25 technical college (Fachschule)</li> </ol>
----------------------	---

	<p>26 university of applied sciences (Fachhochschule)</p> <p>27 university</p> <p>For the ASU source, it has not been possible to differentiate between in-firm and external vocational training since 01 July 2006 due to changes in the procedure. For the XASU source, this differentiation is not possible at all over the entire period of time. In both cases, the mixed category is coded to "in-firm training". For the ASU source, it has neither been possible to differentiate between "no completed vocational training" and "not specified" since 01 July 2006.</p>
Notes on quality	<p>"For certain sub-groups there is a large proportion of missing data because this variable is not of importance as regards social security (this affects people in marginal part-time employment, for example)." (Meinken / Koch 2004, p. 63).</p> <p>Therefore, we advise against a quantitative analysis of the variable for the ASU source from 2006 forward.</p>

### 5.3.5 School-leaving qualification (schbild)

Variable label	school-leaving qualification
Variable name	schbild
Category	personal variables
Origin	ASU, XASU
Data type	numerical
Hierarchy	none
Detailed description	Contains the school-leaving qualification held by the individual at the beginning of the period of unemployment or job search.
Notes on quality	Although the degree of completeness in the XASU has continuously increased over the years, it remains comparatively low (partially under 50 %).

## 5.4 Information on employment, benefit receipt and job-search

### 5.4.1 Reason for notification/ reason for end of benefit receipt/ reason for discontinuation of unemployment benefit II/ reason for deregistration, aggregated (grund\_gr)

Variable label	reason for notification/ reason for end of benefit receipt/ reason for discontinuation of unemployment benefit II/ reason for deregistration, aggregated
Variable name	grund_gr
Category	information on employment, benefit receipt and job-search
Origin	BeH, LeH, (X)LHG, (X)ASU
Data type	numerical
Hierarchy	none
Detailed description	This variable has a different meaning depending on the data source.

### 1) BeH

Reason why the employer submitted the employment notification in question to the social security agencies. Only includes notifications that have information on earnings (i.e. annual, employment interruption and end of employment notifications), initial registrations are not included as they contain no information on earnings. However, this does not involve a loss of information as the details from a registration are transmitted again with the following annual, employment interruption or end of employment notification.

### 2) LeH

Reason for the end of receipt of unemployment benefit, unemployment assistance or maintenance allowance. In contrast, neither the LeH nor the SIAB contains information about the reasons for the start of benefit receipt, as the information for the LeH comes from the notifications submitted by the employment agencies to the health insurance institutions regarding finished periods of benefit receipt.

### 3) LHG/XLHG

Contains the 'reason for discontinuation of Unemployment Benefit II' and indicate the reason why current benefits have been discontinued. The 'reason for discontinuation of Unemployment Benefit II' variable **refers to the individual**, not to the benefit community. If the Unemployment Benefit II receipt of a different member of the benefit community is discontinued, all observations of the members of the benefit community are split on this date, but the reason for discontinuation of Unemployment Benefit II is only available for the individual whose benefit is discontinued. This variable is valid exactly at the end of the original observation.

### 4) ASU/XASU

Contains the deregistration or exit reason. It must be taken into account here that the number of values of the variable has been reduced from 26 April 2003 onwards. For analysis over long periods of time, the old values can be recoded to the currently valid ones using the table below:

old -> new	old -> new	old -> new
29 -> 60	36 -> 61	46 -> 67
30 -> 60	37 -> 66	47 -> 67
31 -> 61	38 -> 66	50 -> 75
32 -> 60	39 -> 71	53 -> 68
33 -> 60	40 -> 69	54 -> 78
34 -> 60	44 -> 74	
35 -> 60	45 -> 77	

This variable is aggregated in the SUF data.

Note on quality	Proportion of valid values (degree of completeness) for the reason for notification in the LHG, XLHG, ASU and XASU data sources:							
		2005	2006	2007	2008	2009	2010	2011
	-----							
	LHG	0.17	0.11	0.19	0.18	0.15	0.12	0.05
	XLHG	0.02	0.06	0.09	0.12	0.17	0.20	0.11
ASU	0.83	0.95	0.98	0.97	0.96	0.89	0.48	
XASU	0.25	0.53	0.52	0.53	0.55	0.49	0.27	

#### 5.4.2 Daily wage, daily benefit rate, aggregated (tentgelt\_gr)

Variable label	daily wage, daily benefit rate, rounded
Variable name	tentgelt_gr
Category	information on employment, benefit receipt and job-search
Origin	BeH, LeH
Data type	numerical
Hierarchy	none
Detailbeschreibung	<p>This variable has a different meaning depending on the data source.</p> <p><b>1) BeH</b></p> <p>Shows the employee's gross daily wage in Euros. Calculated from the fixed-period wages reported by the employer and the duration of the (unsplit) original notification period in calendar days.</p> <p>Earnings exceeding the upper earnings limit for statutory pension insurance were only reported up to this limit. There are two upper earnings limits in the statutory pension insurance scheme: the earnings limit of the pension insurance for wage and salary earners and that of the miners' pension insurance, which is generally higher. However, in the data a differentiation between these two insurance carriers is not possible.</p> <p>Until 1998 employers in principle only reported the earnings which were subject to social security contributions. Earnings below the marginal part-time income threshold were not reported.</p> <p>Since the inclusion of marginal part-time employees in the employment notification procedure on 01 April 1999, earnings below the marginal part-time income threshold have also been recorded; the upper earnings limit still applies as the upper ceiling. In some cases, however, the reported earnings nonetheless exceed the upper earnings limit. Generally, this can probably be attributed to the payment of annual bonuses which the employer can add to the regular earnings in the annual, employment interruption or end of employment notifications. For reasons of data-security, these values are replaced by the upper earnings limit for statutory pension insurance of the according year. Wages above the upper earnings limit for statutory pension insurance can generally be attributed to the payment of annual bonuses which the employer can add to the regular earnings in the annual calculation due to employment interruption or end of employment notifications.</p> <p>In this case, it is irrelevant whether the upper earnings limit in the statutory pension insurance which is decisive for the notification period is</p>

	<p>exceeded as a result of this addition. However, such earnings notifications could also be due to incorrect details in the employment period. (The earnings information, however, may be considered less error-prone due to their insurance relevance.) The marginal part-time income threshold and the upper earnings limit for statutory pension insurance differ from year to year as well as between Eastern and Western Germany (the decisive factor is the location of the establishment). An overview of these limits and thresholds can be found under <a href="http://fdz.iab.de">http://fdz.iab.de</a>.</p> <p>A daily wage reported as 0 euros can be put down to “employment interruption notifications”. During these periods, the employment relationship continues to exist in legal terms, but without pay. This is the case for periods of illness after the end of continued payment of wages, for periods of maternity leave and for sabbaticals.</p> <p>The daily wage is shown with two decimal places. All values greater than 0 and smaller than 0.01 were rounded up to 0.01. This makes it possible to identify the above-mentioned employment interruption notifications with the condition daily wage = 0.</p> <p><b>2) LeH</b></p> <p>Shows the daily benefit rate, converted into euros in each case. For observations with an original start date prior to 01 January 1998 the daily benefit rate applies to working days, while for observations with an original start date from 01 January 1998 onwards it applies to calendar days.</p> <p>Since 01 January 2005, a daily benefit rate reported as 0 euros can be put down to benefit suspension periods or interruptions of benefit payments. If a reason for end of benefit is reported for an observation with a daily benefit rate equal to 0, then it is a notification of interruption of benefit payments. In the case of benefit suspensions (caused by violations of insurance conditions), the entitlement is the same as before the start of the benefit suspension period.</p> <p>The variable is rounded in the SUF data.</p>
--	--

### 5.4.3 Beruf – ausgeübte bzw. letzte Tätigkeit (beruf\_gr)

Variable label	occupation – activity performed / last activity , aggregated
Variable name	Beruf_gr
Category	employment / benefit receipt
Origin	BeH, ASU, XASU
Data type	numerical
Hierarchy	none

Detailed description	<p>The occupational classification consists of a 3-digit code and comprises about 330 values.</p> <p>The three-digit occupational classification is aggregated to 120 occupational groups each consisting of at least 1,000 observations.</p> <p>This variable has a different meaning depending on the data source.</p> <p><b>1) BeH</b></p> <p>Occupational title of the job performed by the employee during the notification period. For this, the employer encodes the employee's job in accordance with the "Classification of Occupations. Systematic and Alphabetical Directory of Job Titles" (published by the Federal Employment Agency, Nuremberg, 1988), which contains approx. 25,000 job titles. The occupational classification consists of a 3-digit code and comprises about 330 values.</p> <p>If more than one job title with different classification codes applies to one employee, the employer is required to select the job title that best defines the main activity performed (see BA 2005, p. V).</p> <p><b>2) ASU/XASU</b></p> <p>For ASU observations, this variable contains the occupation of the last activity performed. For XASU observations, the variable contains the last successfully completed vocational training.</p>
----------------------	--

#### 5.4.4 Occupational status and working hours (stib)

Variable label	occupational status and working hours
Variable name	stib
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Hierarchy	none
Detailed description	<p>Employee's occupational status during the notification period.</p> <p>0 in vocational training  1 unskilled worker  2 skilled worker  3 master craftsman, foreman  4 employee  7 home workers, freelance home workers  8 part-time (not eligible for unemployment benefits)  9 part-time (eligible for unemployment benefits)</p> <p>For the distinction between full-time and part-time employees the decisive factor is the ratio between the contracted hours and the usual working hours in the establishment. The variable actually provides details about the occupational status for full-time employees only, whilst for part-time employees it only records whether their working hours exceed a certain limit or not. This limit was 20 hours of work per week until 1978, between 1979 and 1987 it was 15 hours per week, and</p>

	<p>since 1988 it has been 18 hours per week.</p> <p>The distinction between blue-collar employees in full-time employment and white-collar employees centres solely on the type of pension insurance institution (Federal Social Insurance Office for Salaried Employees – BfA – for white-collar workers, and Land Social Insurance Office – LVA – for blue-collar workers). The "employees in vocational training" category covers not only trainees / apprentices, placement workers and interns but also semi-skilled trainees, students at colleges for health occupations and participants in subsidised further vocational training, retraining and induction training.</p> <p>If more than one code is possible for an employee, the employer is required to classify the job according to the activity which is performed most. If this cannot be ascertained clearly, the code of the higher occupational status is to be entered (see BA 2005, p. VI)</p>
--	--

#### 5.4.5 Employment status (erwstat\_gr)

Variable label	employment status, aggregated
Variable name	Erwstat_gr
Category	personal variables
Origin	BeH, LeH, ASU, XASU
Data type	numerical
Hierarchy	none
Detailed description	<p>This variable has a different meaning depending on the data source.</p> <p><b>1) BeH</b></p> <p>Contains the category of the employment notification, which was introduced along with the new notification procedure (DEÜV) as of 01 January 1999. It indicates contribution- or benefit-related particularities of the employment relationship.</p> <p>If multiple keys apply to an employment notification, the smallest must be indicated by the reporting employer. The majority of these cases are employment relationships subject to social security contributions without any exceptions which are kept under key number 101. Therefore, it is possible that these employees are slightly overestimated. The notification procedure states that changes in the employment status - e.g. when an apprentice is employed with his/her training company after completing his/her vocational training - must be indicated by a new notification (cf. Deutsche BKK 2012, p. 31).</p> <p>The status can be contained in employment notifications that refer to the years prior to 1999 but were not received until 1999 or later. For notifications which were received before 1999, an attempt is made to allocate the notifications to the person groups on the basis of certain rules and with the aid of the 'school education and vocational training', 'occupational status and working hours' and 'occupation' variables as well as other information. In many cases, however, appropriate allocations are not possible. Since 01 April 1999, also employees in marginal part-time employment have been recorded in the DEÜV notification procedure. This group of people can be differentiated via the manifestation 3 (109 and 209 in the original data). For employees in marginal part-time employment, no data prior to the introduction of the notifica-</p>

	<p>tion obligation in 1999 could be collected.</p> <p><b>2) LeH</b>  For LeH observations, the variable employment status contains the grouped benefit type. Thus, it can be differentiated whether a person receives unemployment benefit, unemployment assistance or maintenance allowance or whether contributions to private long-term care insurance are paid by the BA.</p> <p><b>3) ASU / XASU</b>  For ASU observations, the "employment status" variable reports the job search status. Recipients of unemployment benefits (Unemployment Benefit I or II) over the age of 58 who receive benefits under the relaxed conditions according to Section 428 of Social Code Book III (or Section 65 Para. 4 of Social Code Book II) and individuals aged over 58 who are not benefit recipients and are not willing to be placed in employment in the sense of Section 252 Para. 8 Social Code Book VI are recorded as individuals seeking advice.  In XASU observations, the variable employment status reports the "not unemployed, but registered as a job seeker" as well as "unemployed and registered as a job seeker" values.  Since the notification logic might in many cases allow for re-identification of the exact day of birth, the original information on the date were changed by applying the anonymisation procedure described in Chapter 3.4.</p> <p><b>4) LHG/XLHG</b>  Due to the low number of observations, information from the LHG and XLHG data sets is not included in the Scientific Use File.</p>
--	--

#### 5.4.6 Employment before the age of 17 (tage\_jung)

Variable label	days in employment before the lower age limit (year of 17th birthday)
Variable name	tage_jung
Category	information on employment
Origin	BeH
Data type	numerical
Hierarchy	none
Detailed description	This variable reports how many days an individual was employed before the year of his/her 17th birthday. Only the main employment is considered. (Stata syntax: <code>quelle_gr==1 &amp; level2==0</code> ).

#### 5.4.7 Employment after the age of 62 (tage\_alt)

Variable label	days in employment after the upper age limit (year of 62nd birthday)
Variable name	tage_jung
Category	information on employment
Origin	BeH
Data type	numerical

Hierarchy	none
Detailed description	This variable reports how many days an individual is employed after the year of his/her 62nd birthday. Only the main employment is considered. (Stata Syntax: <code>quelle_gr==1 &amp; level2==0</code> ).

## 5.5 Establishment variables

### 5.5.1 Economic activity 93 generated – completed by extrapolation / imputation (w93\_3\_gen)

Variable label	W93_3 Completed by extrapolation / imputation, aggregated
Variable name	W93_3_gen
Category	establishment variables
Origin	BeH
Data type	numerical
Hierarchy	section (1-digit code)   division (2-digit code)   group (3-digit code)   class (4-digit code)   sub-class (5 digit code) of economic activity
Detailed description	<p>Economic activity as a 3-digit code in accordance with the WZ93 classification. From 1998 up to and including 2003, the variable contains the original values from w93_3. Before 1998 and after 2003, the information is either written back / continued or replaced with the help of recoding tables, so that the variable contains time-consistent information on the economic activity based on the economic activity classification WS93. A detailed description can be found in Eberle et al. (2011).</p> <p>For data confidentiality reasons the SIAB-Regional File reports economic activity in 14 aggregated categories. In regions which did not reach the minimum number of observations required for factual anonymity, the economic activities concerned are anonymised and recoded to missing (.z) (see Appendix).</p>

## 5.6 Location data

### 5.6.1 Place of work: region (ao\_region)

Variable label	place of work: district / region
Variable name	ao_region
Category	location data
Origin	BeH
Data type	numerical
Hierarchy	district region   federal state
Detailed description	<p>The variable indicates the district (urban district, rural district or aggregated if the population is lower than 100,000; see Section 3.4.4) in which the employee's establishment is located. The first two digits of the 5-digit district code (Kreisschlüssel) show the code for the federal state (Bundesland), positions 1-3 indicate the regional authority (Regierungsbezirk), and positions 1-5 show the district authority (Kreis). Federal states without a regional authority have a 0 in the third position. Due to the aggregation of the districts, a clear allocation to the regional authority is not possible in a few cases.</p> <p>However, district regions were generated in such a way that an unam-</p>

	<p>biguous allocation to the spatial planning regions (Raumordnungsregionen) is possible.</p> <p>In order to guarantee consistent regional allocations across the entire observation period, the information on the district was recoded to the territorial allocation of 31 December 2010, i.e. in all calendar years, a place of work is assigned to a district in accordance with the boundaries that the district had on 31 December 2010. As the district boundaries have changed over time, cases would occur in which the district code of the location of the establishment would change without the establishment concerned having relocated, if the territorial allocations were not updated.</p> <p>For reasons of data security, the place of work is only given when corresponding to the place of work according to the main spell (level2==0) of an episode. Further parallel places of work are censored (.z).</p>
--	--

### 5.6.2 Commuter status (pendler)

Variable label	commuter status
Variable name	pendler
Category	location data
Origin	BeH, LeH, (X)ASU, (X)LHG
Data type	numerical
Hierarchy	none
Detailed description	<p>The variable 'commuter status' indicates whether an individual's current or most recent place of work corresponds to his or her current place of residence. As a reference for all sources, the current place of work or the place of work from the most recent BeH notification (main employment) is taken as a basis and compared with the current place of residence.</p> <p>The variable contains the following details:</p> <p>0: the district of the place of residence corresponds to the district of the (last available) place of work (no commuter)</p> <p>1: the district of the place of residence does not correspond to the district of the (last available) place of work, but both districts belong to the same district region (intra-regional commuter)</p> <p>2: the district of the place of residence and the district of the (last available) place of work are located in different district regions (inter-regional commuter)</p> <p>This variable is available for:</p> <p>BeH: 1999-2010  LeH: 1999-2010  (X)ASU: 1999 – 2004 und 2007 – 2010  (X)LHG: 2007 – 2010</p> <p>Values for the variable „commuter status“ are only given for main spells (level2==0).</p> <p>If the place of residence for one source is not available in the original data for an individual's episode, the variable 'commuter status' is recoded with .z. .z is also coded if the individual's place of residence is</p>

	available in one source but there is no previous BeH observation with a valid place of work that could be used as a reference.
--	--

## 5.7 Generated technical variables

### 5.7.1 Source of the observation (quelle\_gr)

Variable label	source of the observation, aggregated
Variable name	quelle_gr
Category	generated technical variables
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	numerical
Hierarchy	none
Detailed description	<p>The variable indicates the data source. The original sources ASU and XASU as well as LHG and XLHG were each aggregated to one source for data confidentiality reasons.</p> <p>Hence, it is not possible to distinguish between the types of institution. Because of the adjustments made, the designated time periods of the sources differ from those provided in the SIAB 7510. For further information on data preparation see Chapter 3</p>

### 5.7.2 Observation counter per person (spell)

Variable label	observation counter per person
Variable name	spell
Category	generated technical variables
Origin	BeH, LeH, (X)LHG,(X)ASU
Data type	numerical
Hierarchy	none
Detailed description	<p>The observation counter per person counts a person's observations, beginning with 1. The variable is generated during the episode splitting procedure and refers to the split observations. Using the variable "observation counter per person" it is easy to restore the original sorting order. The observations are sorted first by the start date of the split episode and then by the data source in the following order: BeH, LeH, (X)LHG, (X)ASU. Within the data source BeH, the employment subject to social security with the highest wage comes first, marginal part-time employment is listed last. The data source LeH is sorted according to the benefit type.</p>

## 6 Literatur

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

**Bundesagentur für Arbeit** (Hg.) 2005: Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung. edition January 2005, Nuremberg.

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

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

**Deutsche BKK** (2012): Ratgeber Sozialversicherung 2013, Wolfsburg, 56 S., URL: <http://www.deutschebkk.de/de/arbeitgeber/sozialversicherung/ratgeber-sozialversicherung.html>, (10 December 2013).

**Drews, Nils** 2006: Qualitätsverbesserung der Bildungsvariable in der IAB-Beschäftigtenstichprobe 1975-2001. (FDZ Methodenreport, 05/2006 (de)), Nuremberg, 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)), Nuremberg, 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, volume 126, issue 3, p. 405-436

**Hethey-Maier, Tanja; Seth, Stefan** (2010): The Establishment History Panel (BHP) 1975-2008 \* Handbook Version 1.0.2. (FDZ Datenreport, 04/2010 (en)), Nuremberg, 80 p.

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

**Müller, Walter; Blien, Uwe; Knoche, Peter; Wirth, Heike** (1991): Die faktische Anonymität von Mikrodaten; Band 19 der Schriftenreihe Forum der Bundesstatistik herausgegeben vom Statistischen Bundesamt; Stuttgart.

**vom Berge, Philipp; König, Marion; Seth, Stefan** (2013): Sample of Integrated Labour Market Biographies (SIAB) 1975-2010. (FDZ-Datenreport, 01/2013 (en)), Nürnberg, 65 p.

**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, pp. 468-482.

**Zimmermann, Ralf; Kaimer, Steffen; Oberschachtsiek, Dirk** (2007): Dokumentation des "Scientific Use Files der Integrierten Erwerbsbiographien" (IEBS-SUF V1) Version 1.0. (FDZ Datenreport, 01/2007 (de)), Nuremberg, 62 p.

## 7 Appendix

### 7.1 Overview of date adjustments and spell deletions

	Start and End date observation	Date after censoring	BeH	LeH	(X)ASU	(X)LHG
			1975-2008	1975-2008	2000-2004; 2007-2008	2007-2008
Deletions of spells	Start and end date before 1.1.1975	-	X	X		
	Start and end date between 1.1.2005 and 31.12.2006	-			X	
	Start and end date before 1.1.2000	-			X	
	Start and end date before 1.1.2007	-				X
	Start and end date after 31.12.2008	-		X	X	X
Left-censoring of spells	Start date before 1.1.1975, end date after 1.1.1975	Start date 1.1.1975		X		
	Start date before 1.1.2000, end date after 1.1.2000	Start date 1.1.2000			X	
	Start date between 1.1.2005 and 31.12.2006, end date after 1.1.2007	Start date 1.1.2007			X	X
Right-censoring of spells	Start date before 31.12.2008, end date after 31.12.2008	end date 31.12.2008		X	X	X
	Start date before 31.12.2004, end date between 1.1.2005 and 31.12.2006	end date 31.12.2004			X	

Left- and rightcensoring of spells	Start date before 31.12.2004, end date after 31.12.2006	end date spell 1 31.12.2004; end date spell 2 1.1.2007			X	
------------------------------------	---	--	--	--	---	--

**Table A1:** Outline of date corrections and deletions within the preparations of the SIAB-Regional file 7510

## 7.2 Comparison of the SIAB 7510 und SIAB-Regionalfile 7510

Basic file: number of individuals SIAB 7510		1639325	100.0%	
1	Deletion of observations in the sources BeH and LeH which end before 1975	1639325	100.0%	0
2	Deletion of observations in the sources LeH, (X)ASU and (X)LHG which start after 2010	1636159	99,8 %	-0,2 %
3	Deletion of ASU observations which end before 2000	1635018	99,7 %	-0,1 %
4	Deletion of LHG, XASU and XLHG observations which end before 2007	1633854	99,7 %	-0,0 %
5	Deletion of ASU observations which start after 2004 and end before 2007	1631355	99,5 %	-0,2 %
6	Deletion of observations in all sources which end before the age of 17 or start after the age of 62	1594466	97,3 %	-2,2 %
<b>Number of individuals SIAB-Regional File 7510</b>		<b>1594466</b>	<b>97,3 %</b>	<b>- 2,7%</b>

**Table A2:** Reduction of the number of individuals within the preparations of the SIAB-Regional File 7510

## 7.3 Recoding tables of aggregated variables in comparison with the weakly anonymous version of the SIAB 7510

Data sources, aggregated (quelle_gr)		
SIAB 7508	SIAB-R 7508 (SUF)	Categories
1	1	BeH Employee History File
2	2	LEH Benefit Recipient History File
7, 18	3	(X)ASU Combined Job-Search History Files
16, 19	4	(X)LHG Combined Unemployment Benefit II Recipient History Files

**Table A3:** Recoding of variable "Source of the observation, aggregated" (quelle\_gr)

Employment status (erwstat_gr)		
SIAB 7510	SIAB-R 7510 (SUF)	Categories

101, 140,143	1	Employees liable to social security without special characteristics
102, 141	2	Trainees
109, 209	3	Marginal part-time workers
103, 142	4	Employees in partial retirement
105, 106	5	Interns and student trainees
112, 118, 119, 120, 201, 203, 205	6	Casual workers
1	11	Other employment status
2	12	Unemployment benefit (ALG)
3	13	Unemployment assistance (ALHI)
5	14	Maintenance benefit (UHG)
31	21	Contribution §166b AFG/§207a SGB III
32	22	Unemployed (implying: registered as a job seeker with the BA)
33	23	Illness up to six weeks/incapacitated during unemployment
34	24	Not unemployed, but registered as a job seeker with the BA
35	25	Without status

**Table A4:** Recoding of variable "Source of the observation, aggregated" (quelle\_gr)

<b>Reason for notification/ reason for end of benefit receipt/ reason for discontinuation of unemployment benefit II/ reason for deregistration, aggregated (grund_gr)</b>		
<b>SIAB 7510</b>	<b>SIAB-R 7510 (SUF)</b>	<b>Categories</b>
30, 49 vor 1999 auch: 31,32, 34, 35, 36, 40	1	Deregistration due to end of employment (also to death)
ab 1999: 31	2	Deregistration due to change of health insurance company
ab 1999: 32	3	Deregistration due to change of contribution group
ab 1999: 34, 35	4	Deregistration due to interruption of employment for more than one month (also industrial conflict/dispute)
ab 1999: 36	5	Deregistration due to change of payroll accounting system
ab 1999: 40	6	Simultaneous registration and deregistration due to end of employment
33, 54, 55, 56, 57, 59	7	Deregistration for other reasons
50	8	Annual notification (all employment relationships that exist as of 31 December of the notification year must be declared)
51, 52, 53	9	Employment interruption notification
2002	10	Begin of employment
2001, 2021, 2039, 2041, 2046, 2048, 2051, 2052, 2053, 2056	11	Entitlement to other benefit

2026, 2034, 2038	12	Other reasons
2008, 2015, 2036, 2040, 2042, 2055, 2057, 2058	13	No availability for the labour market
2060	14	End of measure
2023, 2043, 2045, 2059 (zwischen 1980 und 2005)	15	Period of exclusion/shortfall from benefit
2028	16	Entitlement exhausted
2020 (zwischen 1976 und 2005)	17	End of period for maintenance allowance
2003, 2004, 2006, 2007, 2010, 2011, 2012, 2014, 2016, 2017, 2019, 2022, 2024, 2027, 2029, 2031, 2032, 2033, 2035, 2037, 2044, 2047, 2049, 2050, 2054, 2061, 2062, 2063, 2064	18	Other reasons
5031, 5061	20	Placement in ABM, SAM, BSI
5032, 5033, 5034, 5040	21	Placement with integration contract (also disabled), with settling-in allowance and recruitment subsidy for new businesses
5035, 5036, 5079, 5088, 6012, 6027	22	Other measures of active labour market policy
5039, 5071	23	FbW or German language courses
5029, 5037, 5038, 5066, 6001	24	Employment - self sought
5060, 5062, 5063, 5064, 5081, 5086, 5098, 6002, 6003, 6022, 6023, 6024 (ab 1997)	25	Employment (>15 hours) - placed by private job agent or labour administration - found by agency/applicant in job exchange
5030, 5048, 5082, 6005, 6021 (bis 2005)	26	Placement in short-term employment, Employment (non-honorary) – unknown
5046, 5047, 5067, 6006	27	Self-employed
5053, 5068, 6007, 6015, 6025	28	Military/civilian service, military exercise, soldier
5041, 6008	29	Incapacity, Incapacity to work
5042, 5065, 6004	30	Recall / re-employment at the same employer
5056 (bis 2005)	31	Continuance of the current employment
5043, 5070	32	Vocational training
5049, 5069, 6016	33	School education, studies
5045, 5077	34	Relocation
5044, 5074, 6018, 6026	35	Lack of co-operation or availability
5052, 5076	36	Withdrawal from working life
5051	37	Temporary discontinuation of fulfilments
5055, 6028	38	End of need for financial support
5073	39	Non-renewal of notification
5050, 5075, 5087, 6017, 6030, 6031	40	Special regulations
5054, 5078, 5083, 6009, 6010, 6011, 6013, 6014, 6019, 6020, 6029	41	Other reasons
3001, 3003, 4004, 4017, 4018	50	Begin of employment
3006, 4013	53	School education/studies
3007, 4022, 4023	54	Relocation
3008, 4014	55	Self-denotification/local absence without benefit

		receipt
3009, 4006, 4029	56	End of benefit receipt/benefit continuance
3010, 3017, 4001, 4011, 4025	57	Retirement pension, basic financial security in old age
3004, 3016, 4002, 4012	58	(temp.) loss of ability to work
3015, 4016, 4019, 4020, 4021, 4027	59	End of need for financial support
3002, 3005, 3011, 3012, 3013, 4003, 4024, 4026, 4015, 3014, 4005, 4028, 4007, 4008, 4009, 4010	60	Other reasons

**Table A5:** Recoding of variable “Reason for notification/ reason for end of benefit receipt/ reason for discontinuation of unemployment benefit II/ reason for deregistration, aggregated” (grund\_gr)

Occupation – activity performed (beruf_gr)		
SIAB 7510	SIAB-R 7510 (SUF)	Categories
011, 012, 013, 014, 015, 016, 017, 018, 019, 020, 021, 022, 023, 024, 025, 026, 027, 028, 029, 030, 031, 032, 041, 042, 043, 044	1	Farmers until animal keepers and related occupations
051, 052, 053, 054, 055, 056, 057, 058, 059, 060, 061, 062	2	Gardeners, garden workers until forest workers, forest cultivators
071, 072, 073, 074, 075, 076, 077, 078, 079, 080, 081, 082, 083, 084, 085, 086, 087, 088, 089, 090, 091, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112	3	Miners until shaped brick/concrete block makers
121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135	4	Ceramics workers until glass processors, glass finishers
141	5	Chemical plant operatives
142, 143, 144	6	Chemical laboratory workers until vulcanisers
151	7	Plastics processors
161, 162, 163, 164	8	Paper, cellulose makers until other paper products makers
171, 172, 173, 174	9	Type setters, compositors until printers (flat, gravure)
175, 176, 177	10	Special printers, screeners until printers assistants
181, 182, 183, 184	11	Wood preparers until basket and wicker products makers
191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203	12	Iron, metal producers, melters until semi-finished product fettlers and other mould casting occupations
211, 212, 213	13	Sheet metal pressers, drawers, stampers until

		other metal moulders (non-cutting deformation)
221	14	Turners
222, 223, 224	15	Drillers until borers
225, 226	16	Metal grinders until other metal-cutting occupations
231, 232, 233, 234, 235, 242, 243, 244	17	Metal polishers until metal bonders and other metal connectors
241	18	Welders, oxy-acetylene cutters
251, 252, 263	19	Steel smiths until pipe, tubing fitters
261	20	Sheet metal workers
262	21	Plumbers
270, 271, 272	22	Locksmiths, not specified until sheet metal, plastics fitters
273	23	Engine fitters
274, 275	24	Plant fitters, maintenance fitters until steel structure fitters, metal shipbuilders
281	25	Motor vehicle repairers
282, 283, 284	26	Agricultural machinery repairers until precision mechanics
285, 286	27	Other mechanics until watch-, clockmakers
291, 301, 302	28	Toolmakers until precious metal smiths
303, 304, 305, 306	29	Dental technicians until doll makers, model makers, taxidermists
311	30	Electrical fitters, mechanics
312, 313, 315	31	Telecommunications mechanics, craftsmen until radio, sound equipment mechanics
314	32	Electrical appliance fitters
321	33	Electrical appliance, electrical parts assemblers
322	34	Other assemblers
323	35	Metal workers (no further specification)
331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 371, 372, 373, 374, 375, 376, 377, 378	36	Spinners, fibre preparers until skin processing operatives
351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362	37	Cutters until textile finishers
391, 392	38	Bakery goods makers until confectioners (pastry)
401, 402, 403	39	Butchers until fish processing operatives
411, 412	40	Cooks until ready-to-serve meals, fruit, vegetable preservers, preparers
421, 422, 423, 424, , 425, 426, 427, 428, 429, 430, 431, 432, 433	41	Wine coopers until sugar, sweets, ice-cream makers
441, 442	42	Bricklayers until concrete workers
451, 453	43	Carpenters until scaffolders
452	44	Roofers
461, 462	45	Paviors until road makers
463, 464, 465, 466	46	Tracklayers until other civil engineering workers
470, 471, 472	47	Building labourer, general until other building la-

		bourers, building assistants, n.e.c.
481, 482	48	Stucco workers, plasterers, rough casters until insulators, proofers
483, 484, 485, 486	49	Tile setters until screed, terrazzo layers
491, 492, 502, 503, 504	50	Room equippers until other wood and sports equipment makers
501	51	Carpenters
511	52	Painters, lacquerers (construction)
512, 513, 514	53	Goods painters, lacquerers until ceramics/glass painters
521	54	Goods examiners, sorters, n.e.c.
522	55	Packagers, goods receivers, despatchers
531	56	Assistants (no further specification)
541, 542, 543, 544, 545, 546	57	Generator machinists until construction machine attendants
547, 548, 549	58	Machine attendants, machinists helpers until machine setters (no further specification)
601	59	Mechanical, motor engineers
602	60	Electrical engineers
603	61	Architects, civil engineers
604, 605, 606, 607	62	Survey engineers until other engineers
611, 612	63	Chemists, chemical engineers until physicists, physics engineers, mathematicians
621	64	Mechanical engineering technicians
622, 623	65	Electrical engineering technicians until building technicians
624, 625, 626, 627	66	Measurement technicians until remaining manufacturing technicians
628	67	Other technicians
629	68	Foremen, master mechanics
631, 632	69	Biological specialists until physical and mathematical specialists
633, 634	70	Chemical laboratory assistants until photo laboratory assistants
635	71	Technical draughtspersons
681	72	Wholesale and retail trade buyers, buyers
682	73	Salespersons
683, 684, 685, 686	74	Publishing house dealers, booksellers until service-station attendants
687, 688	75	Commercial agents, travellers until mobile traders
691, 692	76	Bank specialists until building society specialists
693, 694	77	Health insurance specialists (not social security) until life, property insurance specialists
701	78	Forwarding business dealers
702, 703, 704, 705, 706	79	Tourism specialists until cash collectors, cashiers, ticket sellers, inspectors
711, 712, 713, 715, 716	80	Railway engine drivers until street attendants
714	81	Motor vehicle drivers
721, 722, 723, 724, 725, 726	82	Navigating ships officers until air transport occu-

		pations
731, 732, 733, 734	83	Post masters until telephonists
741	84	Warehouse managers, warehousemen
742	85	Transportation equipment drivers
743, 744	86	Stowers, furniture packers until stores/transport workers
751	87	Entrepreneurs, managing directors, divisional managers
752, 753	88	Management consultants, organisers until chartered accountants, tax advisers
761, 762, 763	89	Members of Parliament, Ministers, elected officials until association leaders, officials
771, 772	90	Cost accountants, valuers until accountants
773	91	Cashiers
774	92	Data processing specialists
781	93	Office specialists
782, 783	94	Stenographers, shorthand-typists, typists until data typists
784	95	Office auxiliary workers
791, 792	96	Factory guards, detectives until watchmen, custodians
793, 794	97	Doormen, caretakers until domestic and non-domestic servants
801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814	98	Soldiers, border guards, police officers until judicial enforcers
821, 822, 823	99	Journalists until librarians, archivists, museum specialists
831, 832, 833, 834	100	Musicians until scenery/sign painters
835, 836, 837, 838	101	Artistic and assisting occupations (stage, video and audio) until performers, professional sportsmen, auxiliary artistic occupations
841, 842, 843, 844	102	Physicians until Pharmacists
851, 852	103	Non-medical practitioners until masseurs, physiotherapists and related occupations
853	104	Nurses, midwives
854	105	Nursing assistants
855, 857	106	Dietary assistants, pharmaceutical assistants until medical laboratory assistants
856	107	Medical receptionists
861, 863, 891, 892, 893	108	Social workers, care workers until religious care helpers
862	109	Home wardens, social work teachers
864	110	Nursery teachers, child nurses
871, 872, 873, 874	111	University teachers, lecturers at higher technical schools and academies until technical, vocational, factory instructors
875, 876, 877	112	Music teachers, n.e.c. until other teachers
881, 882, 883	113	Economic and social scientists, statisticians until scientists n.e.c.

901, 902	114	Hairdressers until other body care occupations
911, 912	115	Restaurant, inn, bar keepers, hotel proprietors, catering trade dealers until waiters, stewards
913	116	Others attending on guests
921, 922, 923, 924	117	Housekeeping managers until employees by household cheque procedure
931, 932	118	Laundry workers, pressers until textile cleaners, dyers and dry cleaners
933, 934	119	Household cleaners until glass, buildings cleaners
935, 936, 937	120	Street cleaners, refuse disposers until machinery, container cleaners and related occupations

**Table A6:** Recoding of variable "Occupation – activity performed" (beruf\_gr)

<b>Economic activity 93 – group of economic activity 93, aggregated (w93_gen_gr)</b>			
<b>SIAB 7510</b>	<b>SIAB-R 7510 (SUF)</b>	<b>Industry</b>	<b>Anonymisation of industries in regions</b>
011-015, 020, 050, 101-103, 111-112, 120, 131-132, 141-145	1	Agriculture, mining and quarrying	1003, 5314, 6434, 9171, 9563, 9775, 10045, 15002
201-205, 231-233, 251-252	2	production of rubber and plastic products	3256, 5954, 7111, 7339, 9179, 13054, 15002, 15087
241-247	3	Chemical industry	6635, 7140, 7232, 9274, 9471, 9473, 9574, 12061, 16073
271-275, 281-287, 291-297,	4	Facture of structural metal products, mechanical engineering	
300-316, 321-323, 331-335, 341-343, 351-355	5	automotive, production of data processing equipment, electrical and optical Engineering	1051, 3151, 5558, 5916, 9774, 13051
151-160, 171-177, 181-183, 191-193, 211-212, 221-223, 261-268, 361-366, 371-372	6	Consumer goods	
551-555	7	hospitality industry	3102, 3158, 5512, 8425, 9174, 9573
450-455	8	Building Industry	
501-505, 511-517, 521-527	9	Sale, maintenance and repair of motor vehicles and household goods	
601-603, 611-612, 621-623, 631-634, 641-642	10	Transport and communication	9171, 9573
651-652, 660, 671-672, 701-703, 711-714, 721-726, 731-732, 741-748	11	credit and insurance intermediation; Land and Housing, Rentals	
401-403, 410, 900, 911-913, 921-927, 930, 950	12	Public and Personal Services, Household services	
801-804, 851-853	13	Education, social and health-care facilities	

751-753, 990	14	Public administration, social security	5512
--------------	----	--	------

**Table A7:** Recoding of variable “Economic activity 93 – group of economic activity 93, aggregated” (w93\_gen\_gr)

<b>Regionalisation of districts (NUTS 3), territorial allocation 31.12.2010 (ao_region)</b>		
<b>SIAB 7510</b>	<b>SIAB-R 7510 (SUF)</b>	<b>District region</b>
1002	1002	KS Kiel
1003	1003	KS Luebeck
1051	1051	Dithmarschen
1053	1053	Herzogtum Lauenburg
1054	1054	Nordfriesland
1055	1055	Ostholstein
1056	1056	Pinneberg
1057	1057	Ploen
1058, 1004	1058	Rendsburg-Eckernfoerde, KS Neumuenster
1059, 1001	1059	Schleswig-Flensburg, KS Flensburg
1060	1060	Segeberg
1061	1061	Steinburg
1062	1062	Stormarn
2000	2000	KS Hamburg
3101	3101	KS Braunschweig
3102	3102	KS Salzgitter
3151	3151	Gifhorn
3152, 3156	3152	Goettingen
3153	3153	Goslar
3154, 3103	3154	Helmstedt, KS Wolfsburg
3155, 3255	3155	Northeim, Holzminden, Osterode am Harz

3157	3157	Peine
3158	3158	Wolfenbuettel
3241	3241	Region Hannover
3251, 3401	3251	Diepholz, KS Delmenhorst
3252	3252	Hameln-Pyrmont
3254	3254	Hildesheim
3256	3256	Nienburg (Weser)
3257	3257	Schaumburg
3351	3351	Celle
3352, 3461	3352	Cuxhaven, Wesermarsch
3353	3353	Harburg
3355	3355	Lueneburg
3356	3356	Osterholz
3357	3357	Rotenburg (Wuemme)
3358	3358	Soltau-Fallingbostel
3359	3359	Stade
3360, 3354	3360	Luechow-Dannenberg, Uelzen
3361	3361	Verden
3403	3403	KS Oldenburg (Oldenburg)
3404	3404	KS Osnabrueck
3451	3451	Ammerland
3452, 3402	3452	Aurich, KS Emden
3453	3453	Cloppenburg
3454	3454	Emsland
3455, 3405, 3462	3455	Friesland, Wittmund, KS Wilhelmshaven
3456	3456	Grafschaft Bentheim

3457	3457	Leer
3458	3458	Oldenburg
3459	3459	Osnabrueck
3460	3460	Vechta
4011	4011	KS Bremen
4012	4012	KS Bremerhaven
5111	5111	KS Duesseldorf
5112	5112	KS Duisburg
5113	5113	KS Essen
5114	5114	KS Krefeld
5116	5116	KS Moenchengladbach
5117	5117	KS Muelheim an der Ruhr
5119	5119	Ks Oberhausen
5120	5120	Ks Remscheid
5122	5122	KS Solingen
5124	5124	KS Wuppertal
5154	5154	Kleve
5158	5158	Mettmann
5162	5162	Rhein-Kreis Neuss
5166	5166	Viersen
5170	5170	Wesel
5314	5314	KS Bonn
5315	5315	KS Koeln
5316	5316	KS Leverkusen
5334	5334	Städtereion Aachen
5358	5358	Dueren

5362	5362	Rhein-Erft-Kreis
5366	5366	Euskirchen
5370	5370	Heinsberg
5374	5374	Oberbergischer Kreis
5378	5378	Rheinisch-Bergischer Kreis
5382	5382	Rhein-Sieg-Kreis
5512	5512	KS Bottrop
5513	5513	KS Gelsenkirchen
5515	5515	KS Muenster
5554	5554	Borken
5558	5558	Coesfeld
5562	5562	Recklinghausen
5566	5566	Steinfurt
5570	5570	Warendorf
5711	5711	KS Bielefeld
5754	5754	Guetersloh
5758	5758	Herford
5762	5762	Hoexter
5766	5766	Lippe
5770	5770	Minden-Luebbecke
5774	5774	Paderborn
5911	5911	KS Bochum
5913	5913	KS Dortmund
5914	5914	KS Hagen
5915	5915	KS Hamm
5916	5916	KS Herne

5954	5954	Ennepe-Ruhr-Kreis
5958	5958	Hochsauerlandkreis
5962	5962	Maerkischer Kreis
5966	5966	Olpe
5970	5970	Siegen-Wittgenstein
5974	5974	Soest
5978	5978	Unna
6411	6411	KS Darmstadt
6412	6412	KS Frankfurt am Main
6413	6413	KS Offenbach am Main
6414	6414	KS Wiesbaden
6431, 6437	6431	Bergstraße, Odenwaldkreis
6432	6432	Darmstadt-Dieburg
6433	6433	Groß-Gerau
6434	6434	Hochtaunuskreis
6435	6435	Main-Kinzig-Kreis
6436	6436	Main-Taunus-Kreis
6438	6438	Offenbach
6439	6439	Rheingau-Taunus-Kreis
6440	6440	Wetteraukreis
6531	6531	Gießen
6532	6532	Lahn-Dill-Kreis
6533	6533	Limburg-Weilburg
6534	6534	Marburg-Biedenkopf
6535	6535	Vogelsbergkreis
6611	6611	KS Kassel

6631	6631	Fulda
6632	6632	Hersfeld-Rotenburg
6633	6633	Kassel
6634	6634	Schwalm-Eder-Kreis
6635	6635	Waldeck-Frankenberg
6636	6636	Werra-Meißner-Kreis
7111	7111	KS Koblenz
7131	7131	Ahrweiler
7132	7132	Altenkirchen (Westerwald)
7133, 7134	7133	Bad Kreuznach, Birkenfeld
7137	7137	Mayen-Koblenz
7138	7138	Neuwied
7140, 7135	7140	Rhein-Hunsrueck-Kreis, Cochem-Zell
7141	7141	Rhein-Lahn-Kreis
7143	7143	Westerwaldkreis
7211	7211	KS Trier
7231	7231	Bernkastel-Wittlich
7232, 7233	7232	Eifelkreis Bitburg-Pruem, Vulkaneifel
7235	7235	Trier-Saarburg
7314	7314	KS Ludwigshafen am Rhein
7315	7315	Mainz, Stadt
7331, 7319	7331	Alzey-Worms, KS Worms
7332, 7316	7332	Bad Duerkheim, KS Neustadt an der Weinstraße
7333, 7336	7333	Donnersbergkreis, Kusel
7334	7334	Germersheim
7335, 7312	7335	Kaiserslautern, KS Kaiserslautern

7337, 7313	7337	Suedliche Weinstraße, KS Landau in der Pfalz
7338, 7318, 7311	7338	Rhein-Pfalz-Kreis, KS Speyer, KS Frankenthal (Pfalz)
7339	7339	Mainz-Bingen
7340, 7320, 7317	7340	Suedwestpfalz, KS Zweibruecken, KS Pirmasens
8111	8111	KS Stuttgart
8115	8115	Boeblingen
8116	8116	Esslingen
8117	8117	Goeppingen
8118	8118	Ludwigsburg
8119	8119	Rems-Murr-Kreis
8121	8121	KS Heilbronn
8125	8125	Heilbronn
8126	8126	Hohenlohekreis
8127	8127	Schwaebisch Hall
8128	8128	Main-Tauber-Kreis
8135	8135	Heidenheim
8136	8136	Ostalbkreis
8212	8212	KS Karlsruhe
8215	8215	Karlsruhe
8216, 8211	8216	Rastatt, KS Baden-Baden
8221	8221	Heidelberg, Stadt
8222	8222	KS Mannheim
8225	8225	Neckar-Odenwald-Kreis
8226	8226	Rhein-Neckar-Kreis
8231	8231	KS Pforzheim
8235	8235	Calw

8236	8236	Enzkreis
8237	8237	Freudenstadt
8311	8311	KS Freiburg im Breisgau
8315	8315	Breisgau-Hochschwarzwald
8316	8316	Emmendingen
8317	8317	Ortenaukreis
8325	8325	Rottweil
8326	8326	Schwarzwald-Baar-Kreis
8327	8327	Tuttlingen
8335	8335	Konstanz
8336	8336	Loerrach
8337	8337	Waldshut
8415	8415	Reutlingen
8416	8416	Tuebingen
8417	8417	Zollernalbkreis
8421	8421	KS Ulm
8425	8425	Alb-Donau-Kreis
8426	8426	Biberach
8435	8435	Bodenseekreis
8436	8436	Ravensburg
8437	8437	Sigmaringen
9161	9161	KS Ingolstadt
9162	9162	KS Muenchen
9171	9171	Altoetting
9172	9172	Berchtesgadener Land
9173, 9182	9173	Bad Toelz-Wolfratshausen, Miesbach

9174	9174	Dachau
9175	9175	Ebersberg
9176	9176	Eichstaett
9177	9177	Erding
9178	9178	Freising
9179	9179	Fuerstenfeldbruck
9181	9181	Landsberg am Lech
9183	9183	Muehldorf a. Inn
9184	9184	Muenchen
9186, 9185	9186	Pfaffenhofen a.d. Ilm, Neuburg-Schrobenhausen
9187, 9163	9187	Rosenheim, KS Rosenheim
9188	9188	Starnberg
9189	9189	Traunstein
9190, 9180	9190	Weilheim-Schongau, Garmisch-Partenkirchen
9271	9271	Deggendorf
9272, 9276	9272	Freyung-Grafenau, Regen
9273	9273	Kelheim
9274, 9261	9274	Landshut, KS Landshut
9275, 9262	9275	Passau, KS Passau
9277, 9279	9277	Rottal-Inn, Dingolfing-Landau
9278, 9263	9278	Straubing-Bogen, KS Straubing
9362	9362	KS Regensburg
9371, 9361	9371	Amberg-Sulzbach, KS Amberg
9372	9372	Cham
9373	9373	Neumarkt i.d. OPf.
9374, 9363, 9377	9374	Neustadt a.d. Waldnaab, KS Weiden i.d. OPf., Tirschenreuth

9375	9375	Regensburg
9376	9376	Schwandorf
9471, 9461	9471	Bamberg, KS Bamberg
9472, 9462, 9477	9472	Bayreuth, KS Bayreuth, Kulmbach
9473, 9463	9473	Coburg, KS Coburg
9474	9474	Forchheim
9475, 9464, 9479	9475	Hof, KS Hof, Wunsiedel i. Fichtelgebirge
9476, 9478	9476	Kronach, Lichtenfels
9562	9562	Erlangen, Stadt
9563	9563	Fuerth, Stadt
9564	9564	Nuernberg, Stadt
9571, 9561, 9575, 9577	9571	Ansbach, KS Ansbach, Neustadt a.d. Aisch-Bad Windsheim, Weißenburg-Gunzenhausen
9572	9572	Erlangen-Hoechstadt
9573	9573	Fuerth
9574	9574	Nuernberger Land
9576, 9565	9576	Roth, KS Schwabach
9663	9663	Wuerzburg, Stadt
9671, 9661	9671	Aschaffenburg, KS Aschaffenburg
9672	9672	Bad Kissingen
9674, 9673	9674	Haßberge, Rhoen-Grabfeld
9676	9676	Miltenberg
9677	9677	Main-Spessart
9678, 9662	9678	Schweinfurt, KS Schweinfurt
9679, 9675	9679	Wuerzburg, Kitzingen
9761	9761	KS Augsburg
9771	9771	Aichach-Friedberg

9772	9772	Augsburg
9774	9774	Guenzburg
9775	9775	Neu-Ulm
9777, 9762	9777	Ostallgaeu, KS Kaufbeuren
9778, 9764	9778	Unteralldgaeu, KS Memmingen
9779, 9773	9779	Donau-Ries, Dillingen a.d. Donau
9780, 9763, 9776	9780	Oberallgaeu, KS Kempten (Allgäu), Lindau (Bodensee)
10041	10041	Regionalverband Saarbruecken
10042, 10046	10042	Merzig-Wadern, St. Wendel
10043	10043	Neunkirchen
10044	10044	Saarlouis
10045	10045	Saarpfalz-Kreis
11000	11000	KS Berlin
12052	12052	Cottbus, Stadt
12054	12054	Potsdam, Stadt
12060	12060	Barnim
12061	12061	Dahme-Spreewald
12062	12062	Elbe-Elster
12063	12063	Havelland
12064	12064	Maerkisch-Oderland
12065	12065	Oberhavel
12066	12066	Oberspreewald-Lausitz
12067, 12053	12067	Oder-Spree, KS Frankfurt (Oder)
12068, 12070	12068	Ostprignitz-Ruppin, Prignitz
12069, 12051	12069	Potsdam-Mittelmark, KS Brandenburg an der Havel
12071	12071	Spree-Neiße

12072	12072	Teltow-Flaeming
12073	12073	Uckermark
13003	13003	KS Rostock
13051, 13053	13051	Bad Doberan, Guestrow
13052, 13056	13052	Demmin, Mueritz
13054	13054	Ludwigslust
13055, 13002	13055	Mecklenburg-Strelitz, KS Neubrandenburg
13057	13057	Nordvorpommern
13058, 13006	13058	Nordwestmecklenburg, KS Wismar
13059, 13001, 13062	13059	Ostvorpommern, KS Greifswald, Uecker-Randow
13060, 13004	13060	Parchim, KS Schwerin
13061, 13005	13061	Ruegen, KS Stralsund
14511	14511	Chemnitz, Stadt
14521	14521	Erzgebirgskreis
14522	14522	Mittelsachsen
14523	14523	Vogtlandkreis
14524	14524	Zwickau
14612	14612	Dresden, Stadt
14625	14625	Bautzen
14626	14626	Goerlitz
14627	14627	Meißen
14628	14628	Saechsische Schweiz-Osterzgebirge
14713	14713	Leipzig, Stadt
14729	14729	Leipzig
14730	14730	Nordsachsen
15002	15002	Halle (Saale), Stadt

15003	15003	Magdeburg, Stadt
15082, 15001	15082	Anhalt-Bitterfeld, KS Dessau-Roßlau
15083, 15086	15083	Boerde, Jerichower Land
15084	15084	Burgenlandkreis
15085	15085	Harz
15087	15087	Mansfeld-Südharz
15088	15088	Saalekreis
15089	15089	Salzlandkreis
15090, 15081	15090	Stendal, Altmarkkreis Salzwedel
15091	15091	Wittenberg
16051	16051	KS Erfurt
16053	16053	KS Jena
16061	16061	Eichsfeld
16062, 16065	16062	Nordhausen, Kyffhaeuserkreis
16063, 16056	16063	Wartburgkreis, KS Eisenach
16064	16064	Unstrut-Hainich-Kreis
16066, 16054	16066	Schmalkalden-Meiningen, KS Suhl
16067, 16068	16067	Gotha, Soemmerda
16069, 16072	16069	Hildburghausen, Sonneberg
16070	16070	Ilm-Kreis
16071, 16055	16071	Weimarer Land, KS Weimar
16073	16073	Saalfeld-Rudolstadt
16075, 16074	16075	Saale-Orla-Kreis, Saale-Holzland-Kreis
16076, 16052	16076	Greiz, KS Gera
16077	16077	Altenburger Land

**Table A8:** Regionalisation SIAB-Regional file 7510 (ao\_region)

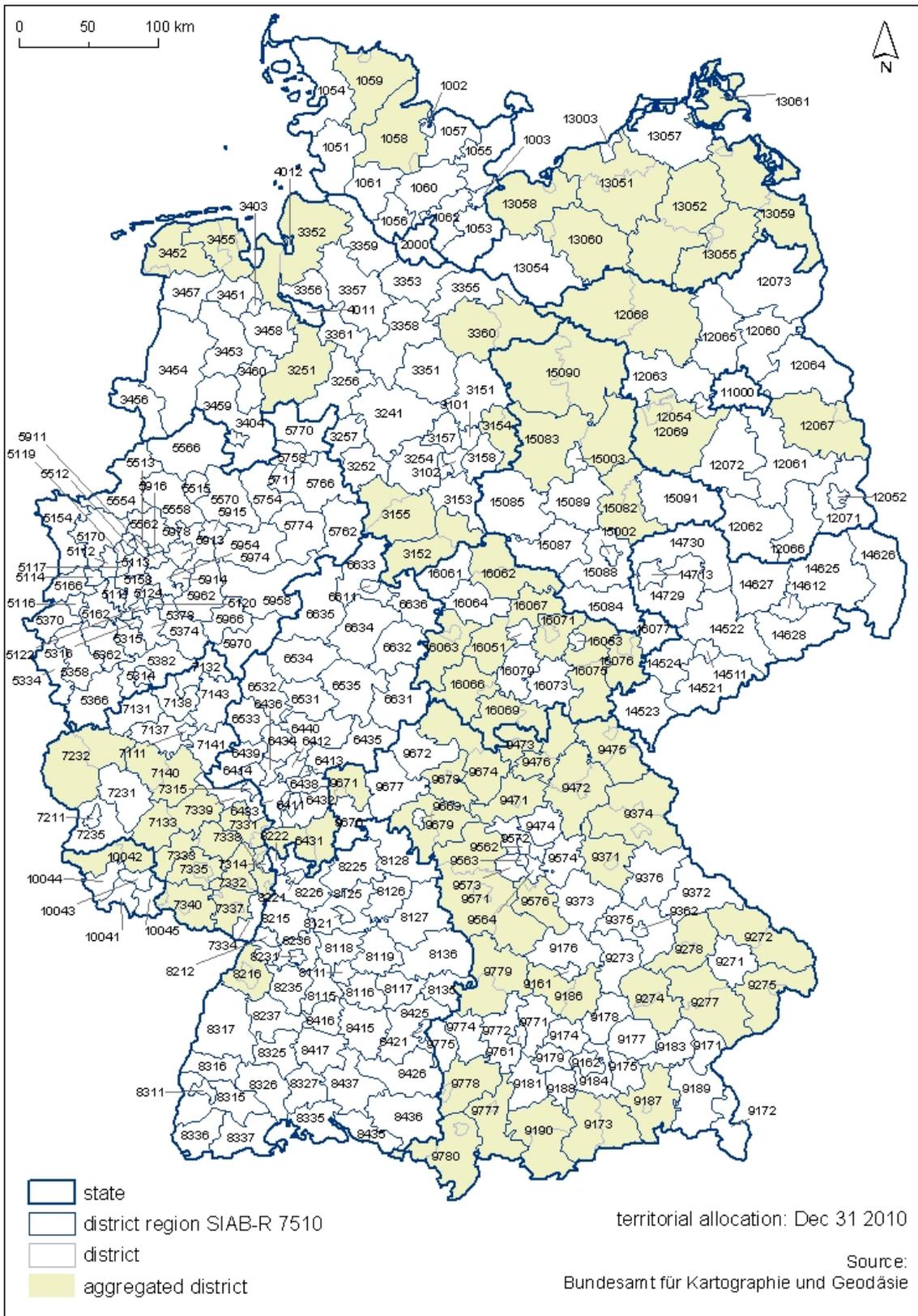


Figure A1: Regionalisation of SIAB Regional 7510 (ao\_region)

## Frequencies

Frequency tables and overviews of the individual values and labels of the variables can be found in separate files under <http://fdz.iab.de/en.aspx>.

## 8 List of abbreviations

<b>AA</b>	Agentur für Arbeit / Arbeitsamt	employment agency / employment office
<b>ALG</b>	Arbeitslosengeld	unemployment benefit
<b>ARGE</b>	Arbeitsgemeinschaft	cooperation of employment agencies and municipalities
<b>ASU</b>	Arbeitsuchende-Historik	Jobseeker History
<b>A2LL</b>	Arbeitslosengeld II – Leistungen zum Lebensunterhalt	unemployment benefit II - benefits to secure a livelihood
<b>BA</b>	Bundesagentur für Arbeit	Federal Employment Agency
<b>BeH</b>	Beschäftigten-Historik	Employee History
<b>BfA</b>	Bundesversicherungsanstalt für Angestellte	Federal Social Insurance Office for Salaried Employees
<b>BHP</b>	Betriebs-Historik-Panel	Establishment History Panel
<b>BMAS</b>	Bundesministerium für Arbeit und Soziales	Federal Ministry of Labour and Social Affairs
<b>coArb</b>	Computerunterstützte Arbeitsvermittlung (operatives Verfahren zur Verwaltung der Vermittlung (Altverfahren))	computer-aided job placement (procedure for the administration of job placements – old procedure)
<b>DEÜV</b>	Verordnung über die Erfassung und Übermittlung von Daten für die Träger der Sozialversicherung – Datenerfassungs- und –übermittlungsverordnung	Data Collection and Transmission Regulation - regulation on the collection and transmission of data for the social security agencies
<b>DEVO</b>	Zweite VO über die Erfassung von Daten für die Träger der Sozialversicherung und für die BA – Datenerfassungs-Verordnung –	Data Collection Regulation - second regulation on the collection of data for the social security agencies and for the Federal Employment Agency
<b>DÜVO</b>	Zweite VO über die Datenübermittlung auf maschinell verwertbaren Datenträgern im Bereich der Sozialversicherung und der BA – Datenübermittlungs-Verordnung –	Data Transmission Regulation - second regulation on the transfer of data on machine-readable data media in the field of social security and the BA
<b>EDV</b>	Elektronische Datenverarbeitung	Electronic data processing
<b>FDZ</b>	Forschungsdatenzentrum	Research Data Centre
<b>FELEG</b>	Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit	Act on the Support in Case of Termination of Farming Activities
<b>gAw</b>	Träger mit getrennter Aufgabenwahr-	Municipalities exercising their duties

	nehmung	separately
<b>gE</b>	Gemeinsame Einrichtung	Joint facility
<b>gT</b>	Getrennte Trägerschaft	Separated responsibilities
<b>IAB</b>	Institut für Arbeitsmarkt- und Berufsforschung	Institute for Employment Research
<b>IABS</b>	IAB-Beschäftigtenstichprobe	IAB Employment Samples
<b>IEB</b>	Integrierte Erwerbsbiographien	Integrated Employment Biographies
<b>IEBS</b>	Stichprobe der Integrierten Erwerbsbiographien	Integrated Employment Biographies Sample
<b>ISIC</b>	International Standard Industrial Classification of All Economic Activities	International Standard Industrial Classification of All Economic Activities
<b>ITM</b>	IT- und Informationsmanagement des Instituts für Arbeitsmarkt- und Berufsforschung	IT and information management of the Institute for Employment Research
<b>LIAB</b>	Linked-Employer-Employee-Daten des IAB	Linked employer-employee data of the IAB
<b>LeH</b>	Leistungsempfänger-Historik	Benefit Recipient History
<b>LHG</b>	Leistungs-Historik Grundsicherung	Unemployment Benefit II Recipient History
<b>LVA</b>	Landesversicherungsanstalt	Land Social Insurance Office
<b>MTH</b>	Maßnahmeteilnehmer-Historik	Participants-in-Measures History File
<b>NACE</b>	Nomenclature générale des activités économiques dans les communautés européennes	Nomenclature générale des activités économiques dans les communautés européennes
<b>Schwbg</b>	Gesetz zur Sicherung der Eingliederung Schwerbehinderter in Arbeit, Beruf und Gesellschaft – Schwerbehindertengesetz-	law to guarantee the integration of persons with severe disabilities into employment and society – Severely Disabled Persons Act
<b>SGB</b>	Sozialgesetzbuch	German Social Code
<b>SIAB</b>	Stichprobe der Integrierten Arbeitsmarktbiographien	Sample of Integrated Labour Market Biographies
<b>VerBIS</b>	Vermittlungs- und Beratungsinformationssysteme	Information System for Placement and Counselling
<b>XASU</b>	Arbeitsuchenden-Historik aus XSozial-BA-SGB II	Jobseeker History from XSozial-BA-SGB II
<b>XLHG</b>	Leistungsempfänger-Historik Grundsicherung aus XSozial-BA-SGB II	Unemployment Benefit II Recipient History from XSozial-BA-SGB II
<b>zKT</b>	Zugelassener kommunaler Träger	Authorised municipalities

## Imprint

FDZ-Datenreport 09/2013\_EN

### Publisher

The Research Data Centre (FDZ)  
of the Federal Employment Agency  
in the Institute for Employment Research  
Regensburger Str. 104  
D-90478 Nuremberg

### Editorial staff

Stefan Bender, Philipp vom Berge

### Technical production

Veronika Zakrocki

### All rights reserved

Reproduction and distribution in any form, also in parts,  
requires the permission of FDZ

### Download

[http://doku.iab.de/fdz/reporte/2013/DR\\_09-13\\_EN.pdf](http://doku.iab.de/fdz/reporte/2013/DR_09-13_EN.pdf)

### Internet

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

### Corresponding author:

Philipp vom Berge  
The Research Data Centre (FDZ)  
Regensburger Str. 104  
90478 Nuremberg  
Germany  
Phone: +49(0)911 / 179-5020  
<mailto:philipp.berge@iab.de>