RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

# FDZ-DATENREPORT

Documentation of labour market data

## **05|2019 EN** Linked Employer-Employee Data from the IAB: LIAB Longitudinal Model (LIAB LM) 1975-2017

Lisa Schmidtlein, Stefan Seth, Matthias Umkehrer

Bundesagentur für Arbeit

# Linked Employer-Employee Data from the IAB: LIAB Longitudinal Model (LIAB LM) 1975-2017

Lisa Schmidtlein, Stefan Seth, Matthias Umkehrer
(Institute for Employment Research (IAB))

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

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

# Contents

# Abstract

This data report describes the LIAB Longitudinal Model 1975 – 2017 (LIAB LM 7517).

# Zusammenfassung

Dieser Datenreport beschreibt das LIAB-Längsschnittmodell 1975 – 2017 (LIAB LM 7517).



# Keywords

German administrative micro data, labour market data, linked employer-employee data, data manual

# Acknowledgements

We would like to thank our colleagues in the Research Data Centre (FDZ), the department Establishments and Employment (D1) as well as the department Data and IT-Management (DIM) of the Institute for Employment Research for their cooperation and support. This Datenreport is a revision of the FDZ-Datenreport 10/2016 (authors: Heining, Klosterhuber, Lehnert, Seth). Additionally, individual passages were adopted from IAB-internal data documentations by IAB-DIM.


# Data availability

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

# 1 Introduction and outline

## 1.1 Introduction

The LIAB Longitudinal Model (LM) 7517 is one of the linked employer-employee datasets from the Institute for Employment Research (IAB) provided for research purposes by the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the IAB. It links information on establishments from the IAB Establishment Panel, an annual establishment survey, with information on individuals employed at those establishments from the process-generated data of the BA.

The IAB Establishment Panel is an annual representative survey on various topics such as the determinants of labour demand. It has been conducted by the IAB since 1993 in West Germany and since 1996 in East Germany. The IAB Establishment Panel is the central basis for the analysis of labour demand in Germany. **This data report focuses on the description of data on individuals in the LIAB LM 7517. More detailed information on the IAB Establishment Panel can be found on the FDZ's websites** (https://fdz.iab.de/en/FDZ_Establishment_Data/IAB_Establishment_Panel.aspx).

Data on individuals come from the Integrated Employment Biographies (IEB) of the IAB. The IEB comprise all individuals showing one of the following statuses at least once during the observation period:

- employment subject to social security (recorded from 1975 onwards)
- marginal part-time employment (recorded from 1999 onwards)
- receipt of benefits in accordance with Social Code Book III (recorded from 1975 onwards) or Social Code Book II (recorded from 2005 onwards)
- registered with the Federal Employment Agency (Bundesagentur für Arbeit - BA) as jobseeker (recorded from 1997 onwards)
- participation in an employment or training measure (recorded from 2000 onwards)

These data originate from different sources. They are merged in the IEB and depict the statuses exact to the day.

Whilst the Employee History (Beschäftigtenhistorik - BeH) is the origin of the information on employment subject to social security and marginal part-time employment, the receipt of benefits in accordance with Social Code Book III (SGB III) and Social Code Book II (SGB II) is recorded in the Benefit Recipient History (Leistungsempfängerhistorik - LeH) and the Unemployment Benefit II Recipient History (Leistungshistoriken Grundsicherung - LHG and XLHG). The Jobseeker Histories (Arbeitsuchendenhistoriken - ASU and XASU) are the data source for the periods of job search recorded by the BA, whilst participation in employment and training measures is recorded in the Participation-in-Measures History File (Maßnahmeteilnahmehistoriken – MTH and XMTH). **Data on individuals in the LIAB LM 7517 only include information from the sources Employee History (BeH) and Benefit Recipient History (LeH). All other sources are not included in the individual data of the LIAB LM 7517**.

The establishments surveyed in the IAB Establishment Panel build the basis for sampling the individual data from the IEB. Figure 1 provides an overview of the different sampling steps (see Chap-

ter 3 for more details). As a starting point, establishments being part of the panel period 2009 – 2016 are selected (so called *panel cases*). This is necessary because of size restrictions. In the next step, all individuals are selected who, according to the process-generated data, where employed at one of those establishments at least once between 2008 and 2017. Finally, the administrative employment biographies of these individuals from the IEB during the period 1975 – 2017 are merged to the LIAB LM 7517.

Despite the restriction concerning the selection of establishments when sampling the individual data, the LIAB LM 7517 includes the survey data of all establishments surveyed in the IAB Establishment Panel between 1993 and 2017 (i.e. also the survey data of establishments that are not *panel cases*). However, note that only for the selected *panel cases* all workers of an establishment can be observed in the individual data. Additionally, the individual data also include workers employed at establishments not being part of the survey. For those establishments, again not all workers can be observed in the individual data. However, it is possible to merge administrative establishment-level information from the BHP for all establishments covered by the LIAB LM 7517 (see https://fdz.iab.de/en/FDZ_Establishment_Data/Establishment_History_Panel.aspx).

The LIAB LM 7517 is produced at the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the IAB. The dataset covers the employment histories of 1.692.611 individuals, and their employment biographies are documented in a total of 56.536.054 lines of data. This Datenreport describes the different variables, which are largely based on the original and uncoarsened data. In order to protect the anonymity of the data subjects, some variables are classified as particularly sensitive and are only made available upon special application (see Section 1.2).

1. **Starting point: Establishment survey (IAB Establishment Panel)**

**IAB Establishment Panel**

2. **Select establishments from the IAB Establishment Panel being part of the panel period 2009 – 2016 (*panel cases*)**

***Panel cases***

3. **Select individuals employed at *panel-case* establishments between 2008 and 2017**

**Individuals employed at *panel-case* establishments**

4. **Merge administrative employment biographies from 1975 – 2017 of individuals selected in step 3**

**Employment biographies**

*Fictitious example:*

| persnr | begepi | endepi | establishment is a panel case? (1=yes) | … |
|---|---|---|---|---|
| 1 | 01/01/1975 | 12/31/1975 | 0 | |
| 1 | 01/01/1976 | 06/30/1976 | 0 | |
| … | | | | |
| 1 | 01/01/2017 | 12/31/2017 | 1 | |

**Figure 1: Sampling structure of the LIAB LM 7517. In the case of a panel establishment, all workers employed at the establishment can be observed. For all other establishments, only a subset of workers can be observed in the individual data. However, administrative establishment information can be merged from the Establishment History Panel (BHP) for all establishments covered by the LIAB LM 7517.**

Other than the cross-sectional model of the LIAB, the longitudinal model includes comprehensive employment biographies of individuals of a subsample of establishments repeatedly interviewed in the IAB Establishment Panel. Therefore, employment biographies can still be tracked in the events of e. g. establishment changes or transitions into unemployment within this subgroup. The cross-sectional model, in turn, contains employment biographies as of the reference date June 30 only. However, for all establishments surveyed in the IAB Establishment Panel the employment biographies of all individuals employed at one of these establishments on June 30 of a year in which a survey took place are included. **An overview of the different LIAB models can be found on the FDZ's websites** (https://fdz.iab.de/en/Integrated_Establishment_and_Individual_Data/LIAB.aspx).

This Datenreport is structured as follows. Besides the introduction, Section 1 contains information on data access as well as an outline of the data, the volume structure and a list of variables. A description of the different data sources can be found in Section 2. Sections 3 and 4 discuss data preparation and data quality, whilst Section 5 describes the individual variables.

## 1.2 Data use

### 1.2.1 Data access

The LIAB data are weakly anonymous data and therefore may only be analysed on-site at the FDZ and via subsequent remote data execution.

In order to be able to use the data, it is necessary to submit an application to the FDZ. The FDZ decides on the approval of the research project on behalf of and, if necessary, in coordination with the Federal Ministry of Labour and Social Affairs (Bundesministerium für Arbeit und Soziales – BMAS). When approval has been granted, a data use agreement is concluded with the researcher's scientific institution. Details on applying for the dataset and possibilities for data processing are available on the FDZ's website.

### 1.2.2 Data management

The LIAB data have a modular structure and include labels both in German and English language.[1] The LIAB data are stored in several files (see Figure 2). A first module contains all waves of the IAB Establishment Panel in separate datasets. These waves include year-specific information on employment trends, business policies and development, company investments, innovations in business, public funding, personnel structure, vocational trainings and training posts, worker inflow and outflow, personnel search, wages and salaries, working hours, further trainings or general data of the establishment. The LIAB LM 7517 contains the full survey data of the IAB Establishment Panel, irrespective of whether an establishment was selected for the sampling of individual data or not.

[1] With the Stata commands `label language de` or `label language en` labels can be switched to English or German, respectively.

A second module, which is henceforth called the Individual File, contains identifiers (artificial individual IDs and establishment IDs), personal characteristics, information on employment and benefit receipt, variables regarding place of residence, and technical variables.

A third module, the Basis Establishment File, contains the establishment number, the year, and variables regarding the place of work and economic activities as well as other establishment information as of the reference date of June 30. These administrative data are available for eastern Germany from 1992 onwards. The Basis Establishment File is extracted from the Establishment History Panel (BHP). The BHP comprises all establishments in Germany with at least one employee liable to social security on June 30 of a year. Since 1999, marginal part-time employees are included in this definition, too. The source of the BHP is the Employee History (BeH) of the IAB. In the BHP, individual data of the BeH is aggregated at the establishment-year level via the establishment numbers. Further information on the BHP can be found on the FDZ's websites (https://fdz.iab.de/en/FDZ_Establishment_Data/Establishment_History_Panel.aspx). Variables marked with an '*' in the list of variables (see Table 4) are included in the Basis Establishment File. Upon reasoned request, further establishment characteristics in variable blocks from the BHP can be provided (see Figure 2). In addition, the extension files Worker flows (inflows/outflows) and Establishment dynamics (entries/exits) can be applied for with a justified application.

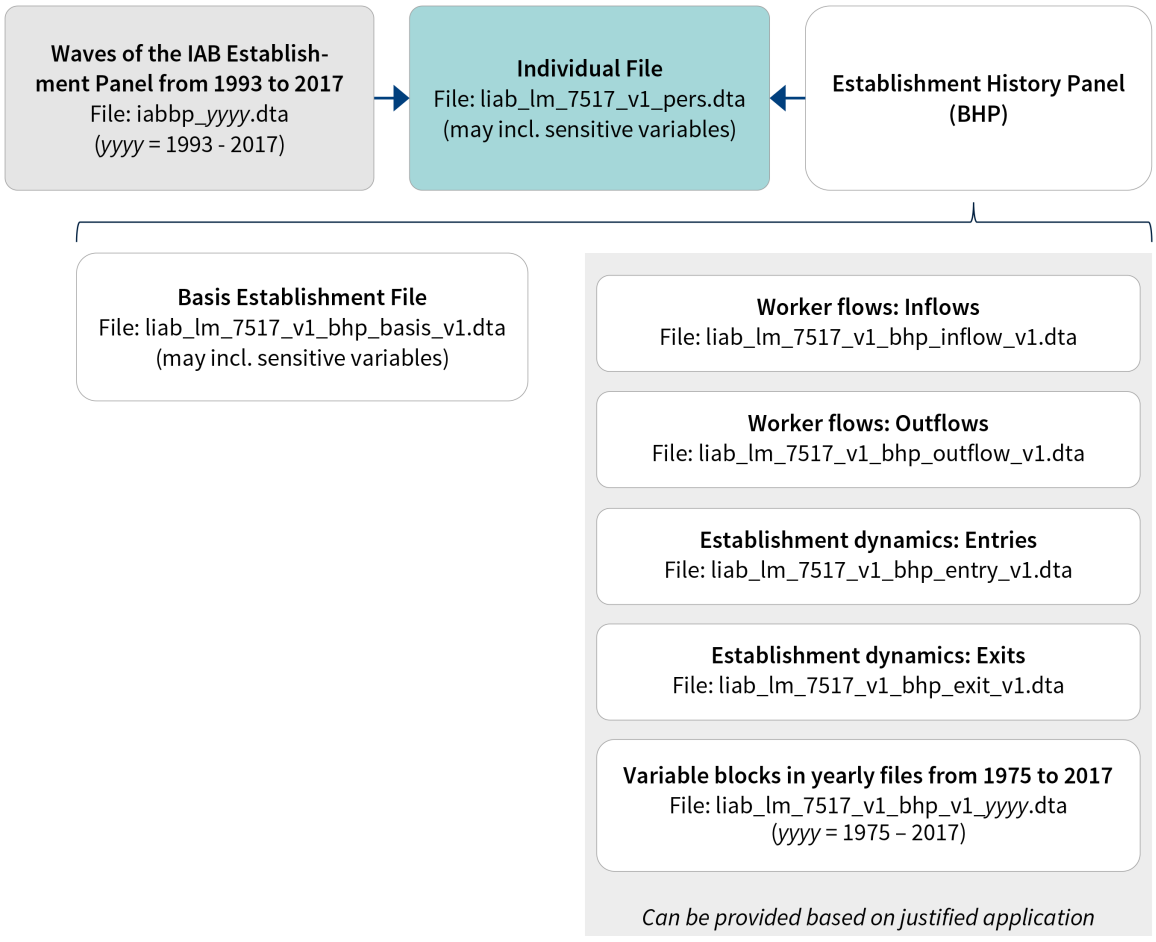



**Figure 2: Data management of the Linked Employer-Employee Data in the LIAB LM 7517**

Establishment variables and individual variables are organised separately, which makes the structure of the data clear and saves storage space. There are different program-specific commands for data formatting and analysis to link the files. For example, linkages in Stata can be conducted via the '`merge`' command. The linkage between the waves of the IAB Establishment Panel and the Individual File is based on the Panel's establishment numbers, which are included in both modules (*idnum*; see Box 1). In contrast, the Basis Establishment File and the Individual File are linked by an artificial establishment number (*betnr*) and year (*jahr*) (see Box 2). The same applies for the extension files of the BHP.

```
use liab_lm_7517_v1_pers.dta
sort idnum
merge m:1 idnum using iabbp_2000.dta
```

**Box 1: Example for Stata 14; Linkage between the Individual File and a single wave of the IAB Establishment Panel**

```
use liab_lm_7517_v1_pers.dta
sort betnr jahr
merge m:1 betnr jahr using LIAB_lm_7517_v1_bhp_basis_v1.dta
```

**Box 2: Example for Stata 14; Linkage between the Individual File and the Basis Establishment File**

Certain variables, which make identification of individuals or establishments easier, are only disclosed in their original form if this is necessary for the analysis objective and is justified explicitly in the application for data access. The following variables are particularly sensitive from a data protection point of view:

Individual File:

- nationality (*nation*)
- month of birth (*gebmon*)
- occupational sub-group (*beruf2010_4*)
- place of residence: employment agency (*wo_aa*)
- place of residence: district (Kreis) (*wo_kreis*)

Basis Establishment File:

- place of work: district (Kreis) (*ao_kreis*)
- economic activity 93 – sub-class of economic activity (five-digit code) (*w93_5*)
- economic activity 03 - sub-class of economic activity (five-digit code) (*w03_5*)

- economic activity 08 - sub-class of economic activity (five-digit code) (*w08_5*)
- date of first appearance establishment number (*grd_dat*)
- date of last appearance establishment number (*lzt_dat*)

All sensitive variables listed in this section are located in the corresponding files, as long as the files have been requested and the request has been approved (see Figure 2).

### 1.2.3 File names of the LIAB LM 7517

Waves of the IAB Establishment Panel

iabbp_*yyyy*.dta, *yyyy* = 1993 – 2017

Individual File

liab_lm_7517_v1_pers.dta

Basis Establishment File

liab_lm_7517_v1_bhp_basis_v1.dta

Extension Files BHP

Stocks with variable blocks

liab_lm_7517_v1_bhp_v1_ *yyyy*.dta, *yyyy* = 1975 – 2017

Worker flows

liab_lm_7517_v1_bhp_inflow_v1.dta

liab_lm_7517_v1_bhp_outflow_v1.dta

Establishment dynamics

liab_lm_7517_v1_bhp_entry_v1.dta

liab_lm_7517_v1_bhp_exit_v1.dta

## 1.3 Changes as compared to LIAB LM 9314

### 1.3.1 Observation period

The LIAB LM 7517 differs from the LIAB LM 9314 data in terms of time periods covered by both the individual and establishment data.

### 1.3.2 Sampling procedure

A more recent panel period was chosen for selecting the *panel cases* from the IAB Establishment Panel. While the LIAB LM 9314 was based on the panel period 2003 – 2011, the LIAB LM 7517 uses

the panel period 2009 – 2016. Additionally, the time period covering the individual employment biographies changes, from 1993 – 2014 to 1975 – 2017 (see Section 3.3).

### 1.3.3 Correction of entries in the municipality of Lahn

For the years 1975 – 1977 there was a misclassification of the establishments and employees in the municipality of Lahn (Lower Saxony) in the BeH. These misclassifications were corrected as far as possible. Further information is available in Section 4.2.

### 1.3.4 Set of variables

The LIAB LM 7517 offers a number of additional variables as compared to the LIAB LM 9314. The variable 'Occupational status and working hours' (*stib*), which was removed in the previous version, is included again in the LIAB LM 7517. However, the restriction applies that the variable is only filled for notifications that were submitted before the introduction of the new occupation code (see Section 4.2). In addition, the variable 'Vocational training (imputed)' (*ausbildung_imp*, see Section 5.4.9) has been added. This variable contains imputed data for observations with originally missing information on the level of vocational training. Additionally, the LIAB LM 7517 now also contains the sensitive variable 'Month of birth' (*gebmon*, see Section 5.4.3) as separate variable. Furthermore, a time-consistent version of the WZ08 industry classification was generated by extrapolation and imputation. Further information can be found in Sections 5.7.12 and 5.7.13. The variables 'Start date of unemployment' (*alo_beg*) and 'Duration of unemployment' (*alo_dau*) are new in the LIAB LM 7517 to provide additional information on individual unemployment (see Sections 5.5.14 and 5.5.15). These variables are already created in the IEB because the LIAB LM 7517 lacks some data sources required for constructing them.

The LIAB LM7517 covers individual employment biographies from 1975 onwards. The previous version capped those biographies before 1993. To provide at least some information on the biographies before 1993, the following biographical variables were generated in the previous version:

- First day in employment (*ein_erw*)
- Number of days in employment (*tage_erw*)
- First day in establishment (*ein_bet*)
- Number of days in establishment (*tage_bet*)
- First day in job (*ein_job*)
- Numbers of days in job (*tage_job*)
- Number of days of benefit receipt (*tage_lst*)

Researchers can now generate these variables themselves, which is why they are no longer included in the LIAB LM 7517. Eberle and Schmucker (2019) offer the corresponding programmes for Stata.

Table 1 provides an overview of the changes in the set of variables across all sources.

**Table 1: Changes in the set of variables**

| Variable | Type of modification | Explanation |
|---|---|---|
| Vocational training (imputed) (*ausbildung_imp*) | N | Newly included, yet with fewer categories than the variable 'Vocational training' (*ausbildung*) |
| Month of birth (*gebmon*) | N | Sensitive variable |
| Occupational status and working hours (*stib*) | N | Included again, yet only filled for notifications submitted before the introduction of the new occupation code |
| w08_3 completed by extrapolation/imputation (*w08_3_gen*) | N | Newly included in Basis Establishment File |
| Art der Vervollständigung w08_3 (*group_w08_3*) | N | Newly included in Basis Establishment File |
| Start date of unemployment (*alo_beg*) | N | The variable is generated within the IEB using multiple data sources. Therefore, it cannot be recreated exactly with the individual data of the LIAB. |
| Duration of unemployment (*alo_dau*) | N | The variable is generated within the IEB using multiple data sources. Therefore, it cannot be recreated exactly with the individual data of the LIAB. |
| Date of birth (*gebdat*) | D | Instead of in a combined variable (*gebdat*), the information on the year and month of birth is now stored in two variables (*gebmon*, *gebjahr*). |
| First day in employment (*ein_erw*) | D | Researchers can now generate this variable themselves using biographies from 1975 onwards. |
| Number of days in employment (*tage_erw*) | D | Researchers can now generate this variable themselves using biographies from 1975 onwards. |
| First day in establishment (*ein_bet*) | D | Researchers can now generate this variable themselves using biographies from 1975 onwards. |
| Number of days in establishment (*tage_bet*) | D | Researchers can now generate this variable themselves using biographies from 1975 onwards. |
| First day in job (*ein_job*) | D | Researchers can now generate this variable themselves using biographies from 1975 onwards. |
| Numbers of days in job (*tage_job*) | D | Researchers can now generate this variable themselves using biographies from 1975 onwards. |
| Number of days of benefit receipt (*tage_lst*) | D | Researchers can now generate this variable themselves using biographies from 1975 onwards. |

N = new, D = dropped

## 1.4 Outline

**Table 2: Outline**

| | |
|---|---|
| Topics/ groups of variables | **Employee History (BeH):**<br>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.<br>**Benefit Recipient History (LeH):**<br>Information on benefit receipt in accordance with Social Code Book III (SGB III) for recipients of unemployment benefit, unemployment assistance and maintenance allowance.<br><br>**IAB Establishment Panel:**<br>Information from the annual Establishment Panel survey with shifting priorities. |
| Data units | Establishments, Employees covered by social security (including marginal part-time employees from 1999 onwards), benefit recipients |
| Number of cases | Linked establishments: 2.187 to 12.261 per year, 41.777 in total<br>Individuals: 135.985 to 1.399.924 per year, 1.692.611 in total<br>Non-overlapping intervals (after episode splitting): 56.586.054 |
| Period covered | The period covered depends on the data source.<br>BeH January 1, 1975 – December 31, 2017<br>(2014: 36-Months-File<br>2015 and 2016: 18-Months-File,<br>2017: 6-Months-File)<br>LeH January 1, 1975 – December 31, 2017 |
| Time reference | Establishments: Referenced to June 30<br>Individuals: Exact to the day |
| Regional structure | German federal states (Bundesländer), districts (Kreise) |
| Date of territorial allocation | Territorial allocation updated as of December 31, 2017 |
| Data collection method | Linked employer-employee data; selection of particular establishments from establishment survey and linkage with their employees' employment biographies from process data |
| Institutions involved | Social security agencies, Federal Employment Agency (Bundesagentur für Arbeit), Kantar Public (formerly TNS Infratest Sozialforschung GmbH) in cooperation with SÖSTRA (Institute for Socio-economic Structural Analysis). |
| Frequency of data collection | Administrative data: continuous<br>IAB Establishment Panel: annually |
| File format and size | Stata<br>Establishment Files (without sensitive variables):<br>IAB Establishment Panel: about 1.7 to 11 MB per year<br>Basis Establishment File: about 455 MB<br>Individual File (without sensitive variables): about 3.5 GB |

| | |
|---|---|
| File architecture | The LIAB data on individuals are stored in one single file. Another file with establishment information taken from the BHP (the Basis Establishment File) is stored in long format and can be linked to the data on individuals by using the variables *betnr* and *year*. Upon reasoned request, further information on establishments from the BHP can be provided.<br>The IAB Establishment Panel files remain unchanged as one file per wave and can be linked to the data on individuals by using the variable *idnum*. |
| Data access | On-site access and subsequent remote data execution |
| Degree of anonymisation | Weakly anonymous |
| Sensitive variables | Individual file:<br>Nationality (*nation*), Month of birth (*gebmon*), Occupational sub-group (*beruf2010_4*), Place of residence – district (*wo_kreis*), Place of residence – employment agency area (*wo_aa*)<br><br>Basis Establishment File:<br>Place of work – district (*ao_kreis*), Economic activity - sub-class of economic activity (5-digit code) (*w93_5*), Economic activity - sub-class of economic activity (5-digit code) (*w03_5*), Economic activity - sub-class of economic activity (5-digit code) (*w08_5*), First appearance of establishment (*grd_dat*), Last appearance of establishment (*lzt_dat*) |
| Citation of data and data documentation | Dataset:<br>Linked Employer-Employee Data from the IAB, Longitudinal Model 1975-2017 (LIAB LM 7517), Nuremberg 2019<br>Data:<br>"The data basis of this paper is the Longitudinal Model 1975 – 2017 of the Linked Employer-Employee Data from the IAB. The data were accessed on-site at the Research Data Centre of the Federal Employment Agency at the Institute for Employment Research (FDZ) and via remote data execution at the FDZ."<br>DOI: 10.5164/IAB.LIABLM7517.de.en.v1<br><br>Data documentation:<br>Schmidtlein, Lisa; Seth, Stefan; Umkehrer, Matthias (2019): Linked Employer-Employee Data from the IAB: LIAB Longitudinal Model (LIAB LM) 1975 – 2017. FDZ-Datenreport, 05/2019 (en), Nuremberg.<br>DOI: 10.5164/IAB.FDZD.1905.en.v1 |
| Dataset version | Linked Employer-Employee Data from the IAB, Longitudinal Model 1975 – 2017 (LIAB LM 7517) – Version 7517 v1;<br>DOI: 10.5164/IAB.LIABLM7517.de.en.v1 |

## 1.5 List of variables

The overview of variables in Table 4 lists the variable names and the longer descriptions of variables. It also provides an overview of whether and how well variables are filled in the individual data sources. Table 3 illustrates the meaning of the shading, which indicates the degree of completeness per variable and source in Table 4.

Example: The variable 'Occupation – current/most recent (KldB 1988)' (*beruf*) is only available for BeH observations. The observations of the data source LeH contain '.n' in this variable.

**Table 3: Degrees of completeness of variables**

| | |
|---|---|
| **h** | Variable is available for the data source.<br>Degree of completeness > 0.85 |
| **m** | Variable is available for the data source.<br>Lower or varying degree of completeness, see description of variable and frequency count |
| **l** | Variable is not available for this data source.<br>Degree of completeness < 0.05 |

Another characteristic is that some variables have a different content depending on the data source. For instance, for BeH observations the variable 'Employment status' (*erwstat*) contains the person group of the employment notification procedure and for LeH observations it contains the type of benefit. In some cases, these differences are not immediately obvious from the variable name.

**Table 4: List of variables with degree of completeness**

| **Variable list** | In Basis Establish-ment File | Page | BeH | LeH |
|---|---|---|---|---|
| **Identifiers** | | 31 | | |
| Individual ID (*persnr*) | | 31 | **h** | **h** |
| Establishment ID (*betnr*) | * | 31 | **h** | **l** |
| Establishment-ID acc. to Survey (*idnum*) | | 32 | **m** | **l** |
| **Generated technical variables** | | 33 | | |
| Counter per person (*spell*) | | 33 | **h** | **h** |
| Source of spell (*quelle*) | | 33 | **h** | **h** |
| Year (*jahr*) | * | 33 | **h** | **h** |
| Status of establishment number (*betr_st*) | | 33 | **h** | **l** |
| **Period of validity** | | 34 | | |
| Original start date (*begorig*) | | 34 | **h** | **h** |
| Original end date (*endorig*) | | 34 | **h** | **h** |
| Episode start date (*begepi*) | | 35 | **h** | **h** |
| Episode end date (*endepi*) | | 35 | **h** | **h** |
| **Personal information** | | 35 | | |
| Gender (*frau*) | | 35 | **h** | **h** |
| Year of birth (*gebjahr*) | | 35 | **h** | **h** |
| Month of birth (*gebmon*) | | 36 | **h** | **h** |
| Nationality (*nation*) | | 36 | **h** | **h** |
| Nationality, grouped (*nation_gr*) | | 36 | **h** | **h** |

| | | | | |
|---|---|---|---|---|
| Marital status (*famst*) | | 37 | l | h |
| Number of children (*kind*) | | 37 | l | h |
| Vocational training (*ausbildung*) | | 37 | m | l |
| Vocational training (imputed) (*ausbildung_imp*) | | 38 | h | l |
| School leaving qualification (*schule*) | | 38 | m | l |
| **Information on employment and benefit receipt** | | 39 | | |
| Daily wage/daily benefit (*tentgelt*) | | 39 | h | h |
| Occupation - current/most recent (KldB 1988) (*beruf*) | | 40 | h | l |
| Occupational group - current/most recent (KldB 2010), 3-digit (*beruf2010_3*) | | 41 | h | m |
| Occupational sub-group - current/most recent (KldB 2010), 4-digit (*beruf2010_4*) | | 41 | h | m |
| Level of requirement - current/most recent job (KldB 2010) (*niveau*) | | 42 | h | m |
| Part-time (*teilzeit*) | | 43 | h | l |
| Occupational status and working hours (*stib*) | | 43 | h | l |
| Employment status (*erwstat*) | | 44 | h | h |
| Transition zone (*gleitz*) | | 45 | h | l |
| Temporary agency work (*leih*) | | 45 | h | l |
| Fixed-term contract (*befrist*) | | 45 | h | l |
| Reason of cancellation/notification/termination (*grund*) | | 46 | h | h |
| Residual claim/planned duration (*restanspruch*) | | 46 | l | h |
| Start date of unemployment (*alo_beg*) | | 47 | m | m |
| Duration of unemployment (*alo_dau*) | | 47 | h | h |
| **Location data** | | 48 | | |
| Place of residence - district (Kreis) (*wo_kreis*) | | 48 | h | h |
| Place of residence - federal state (Bundesland) (*wo_bula*) | | 49 | h | h |
| Place of residence - employment agency (Arbeitsagentur) (*wo_aa*) | | 49 | h | h |
| Place of residence - regional directorate (Regionaldirektion) (*wo_rd*) | | 50 | h | h |
| **Establishment variables** | | 50 | | |
| Classification of economic activities 73 (*w73_3*) | * | 50 | h | l |
| Classification of economic activities 93, subclasses (*w93_5*) | * | 51 | h | l |
| Classification of economic activities 93, groups (*w93_3*) | * | 51 | h | l |
| Classification of economic activities 03, subclasses (*w03_5*) | * | 52 | h | l |

| | | | | |
|---|---|---|---|---|
| Classification of economic activities 03, groups (*w03_3*) | * | 52 | h | l |
| Classification of economic activities 08, sub-classes (*w08_5*) | * | 52 | h | l |
| Classification of economic activities 08, groups (*w08_3*) | * | 53 | h | l |
| w73_3 completed by extrapolation/imputation (*w73_3_gen*) | * | 53 | h | l |
| Type of imputation w73_3 (*group_w73_3*) | * | 54 | h | l |
| w93_3 completed by extrapolation/imputation (*w93_3_gen*) | * | 54 | h | l |
| Type of imputation w93_3 (*group_w93_3*) | * | 54 | h | l |
| w08_3 completed by extrapolation/imputation (*w08_3_gen*) | * | 54 | h | l |
| Type of imputation w08_3 (*group_w08_3*) | * | 55 | h | l |
| Year of first appearance (*grd_jahr*) | * | 55 | h | l |
| First appearance (*grd_dat*) | * | 55 | h | l |
| Year of last appearance (*lzt_jahr*) | * | 56 | h | l |
| Last appearance (*lzt_dat*) | * | 56 | h | l |
| Total number of employees (*az_ges*) | * | 57 | h | l |
| Number of full-time employees (regular workers + others) (*az_vz*) | * | 57 | h | l |
| Number of employees in marginal part-time employment (*az_gf*) | * | 57 | h | l |
| Mean imputed wage all full-time employees (*te_imp_mw*) | * | 58 | h | l |
| Place of work - district (Kreis) (*ao_kreis*) | * | 58 | h | l |
| Place of work - federal state (Bundesland) (*ao_bula*) | * | 59 | h | l |

## 1.6 Volume structure

Column (2) of Table 5 indicates the number of individuals per year in the LIAB LM 7517. In column (3), the establishment numbers (*betnr*) per year, contained in the Individual File, are shown.

In contrast, columns (4) to (7) contain the total number of establishments from the IAB Establishment Panel that can be linked to the individual data. Column (4) displays the number of linked establishments in total. The number of linked establishments according to the variable 'Status of establishment number' (*betr_st*, see Section 5.2.4) is shown by columns (5) to (7).

Thereby, column (5) contains all establishments, that have been originally selected for the LIAB LM 7517 (see also Section 3.3). These so-called *panel cases* have been selected from the IAB Establishment Panel for the period 2009 – 2016. The displayed numbers refer to linkages of establishments which

- have been successfully interviewed (on June 30 of the respective year)
- and for which the observed episode includes this day of reference.

In contrast, column (6) presents the number of linked establishments which

- have been successfully interviewed (on June 30 of the respective year)
- but for which the observed episode excludes this day of reference.

Finally, column (7) shows the number of establishments from the IAB Establishment Panel which are not classified as *panel cases* in the survey for the period 2009 to 2016 but which can be linked to the individual data though.

**Table 5: Volume structure**

| Year | Individuals | Establish-ments | Establishments from the IAB Establishment Panel linked to the individual data | | | |
|---|---|---|---|---|---|---|
| | Total | Total | Total | Establishment selected for LIAB model | | Establishment not selected for LIAB model |
| | | | | *betr_st* == 1 | *betr_st* == 2 | *betr_st* == 3 |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| 1975 | 135 985 | 70 735 | | | | |
| 1976 | 150 451 | 78 959 | | | | |
| 1977 | 167 835 | 84 347 | | | | |
| 1978 | 182 942 | 89 467 | | | | |
| 1979 | 203 935 | 97 216 | | | | |
| 1980 | 222 369 | 102 529 | | | | |
| 1981 | 239 335 | 104 380 | | | | |
| 1982 | 254 273 | 104 940 | | | | |
| 1983 | 268 892 | 108 173 | | | | |
| 1984 | 288 445 | 113 315 | | | | |
| 1985 | 305 704 | 115 323 | | | | |
| 1986 | 329 271 | 119 919 | | | | |
| 1987 | 348 979 | 123 477 | | | | |
| 1988 | 366 464 | 126 622 | | | | |
| 1989 | 390 956 | 132 856 | | | | |
| 1990 | 424 651 | 143 858 | | | | |
| 1991 | 571 840 | 184 054 | | | | |
| 1992 | 671 553 | 212 978 | | | | |
| 1993 | 686 299 | 219 461 | 2 432 | 172 | 126 | 2 258 |
| 1994 | 702 476 | 226 405 | 2 293 | 185 | 123 | 2 104 |
| 1995 | 728 138 | 230 554 | 2 187 | 204 | 143 | 1 979 |
| 1996 | 749 684 | 232 710 | 4 958 | 698 | 466 | 4 250 |
| 1997 | 777 219 | 236 379 | 4 643 | 790 | 534 | 3 843 |
| 1998 | 816 908 | 241 145 | 5 100 | 970 | 662 | 4 114 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 1999 | 879 748 | 306 509 | 5 592 | 1 254 | 899 | 4 322 |
| 2000 | 953 230 | 321 842 | 7 795 | 1 740 | 1 265 | 6 016 |
| 2001 | 988 326 | 319 758 | 8 680 | 2 052 | 1 525 | 6 588 |
| 2002 | 1 015 304 | 310 463 | 8 412 | 2 368 | 1 855 | 6 012 |
| 2003 | 1 047 107 | 315 020 | 8 535 | 2 986 | 2 530 | 5 508 |
| 2004 | 1 084 647 | 323 266 | 8 863 | 3 321 | 2 652 | 5 499 |
| 2005 | 1 085 709 | 324 397 | 9 062 | 3 710 | 2 939 | 5 310 |
| 2006 | 1 127 706 | 334 092 | 9 076 | 4 213 | 3 431 | 4 826 |
| 2007 | 1 190 587 | 342 093 | 9 354 | 4 806 | 4 012 | 4 509 |
| 2008 | 1 257 950 | 348 111 | 9 789 | 5 494 | 4 701 | 4 250 |
| 2009 | 1 275 709 | 358 149 | 10 204 | 6 176 | 5 359 | 3 938 |
| 2010 | 1 299 344 | 370 501 | 10 269 | 6 460 | 5 630 | 3 628 |
| 2011 | 1 339 255 | 382 583 | 10 388 | 6 562 | 5 701 | 3 646 |
| 2012 | 1 357 859 | 386 260 | 10 911 | 6 750 | 5 730 | 3 966 |
| 2013 | 1 374 278 | 388 351 | 11 180 | 7 049 | 6 048 | 3 920 |
| 2014 | 1 389 022 | 392 233 | 11 348 | 7 483 | 6 363 | 3 646 |
| 2015 | 1 396 878 | 392 862 | 11 823 | 7 964 | 6 693 | 3 622 |
| 2016 | 1 399 924 | 393 460 | 12 261 | 8 672 | 7 345 | 3 303 |
| 2017 | 1 393 830 | 391 475 | 11 136 | 7 304 | 6 152 | 3 578 |

**Note:** Values in columns (5), (6) and (7) do not add up to values in column (4).

# 2 Data sources

The administrative individual data were drawn from the Integrated Employment Biographies (IEB) of the IAB. The IEB combine data from different data sources, each of which may contain information from different administrative procedures. In addition, some supplementary variables from these data sources, which are not part of the IEB, are incorporated into the administrative individual data. Figure 3 illustrates the data flows that lead to the LIAB LM 7517 and their relationship to other FDZ data products.

In the individual data of the LIAB LM 7517, only the datasets from the sources Employee History (BeH) and Benefit Recipient History (LeH) are adopted from the IEB.

## 2.1 Employee History (BeH)

The source of data regarding employment is the Employee History (Beschäftigtenhistorik - BeH) of the IAB. The data basis is the integrated notification procedure for health, pension and unemployment insurance, which came into effect as of January 1, 1973 (and was extended to cover East Germany as of January 1, 1991) and is known by the abbreviation DEÜV (previously DEVO / DÜVO) (for further details see: Bender et al. 1996, p. 4 et seq.; 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 at least once a year.

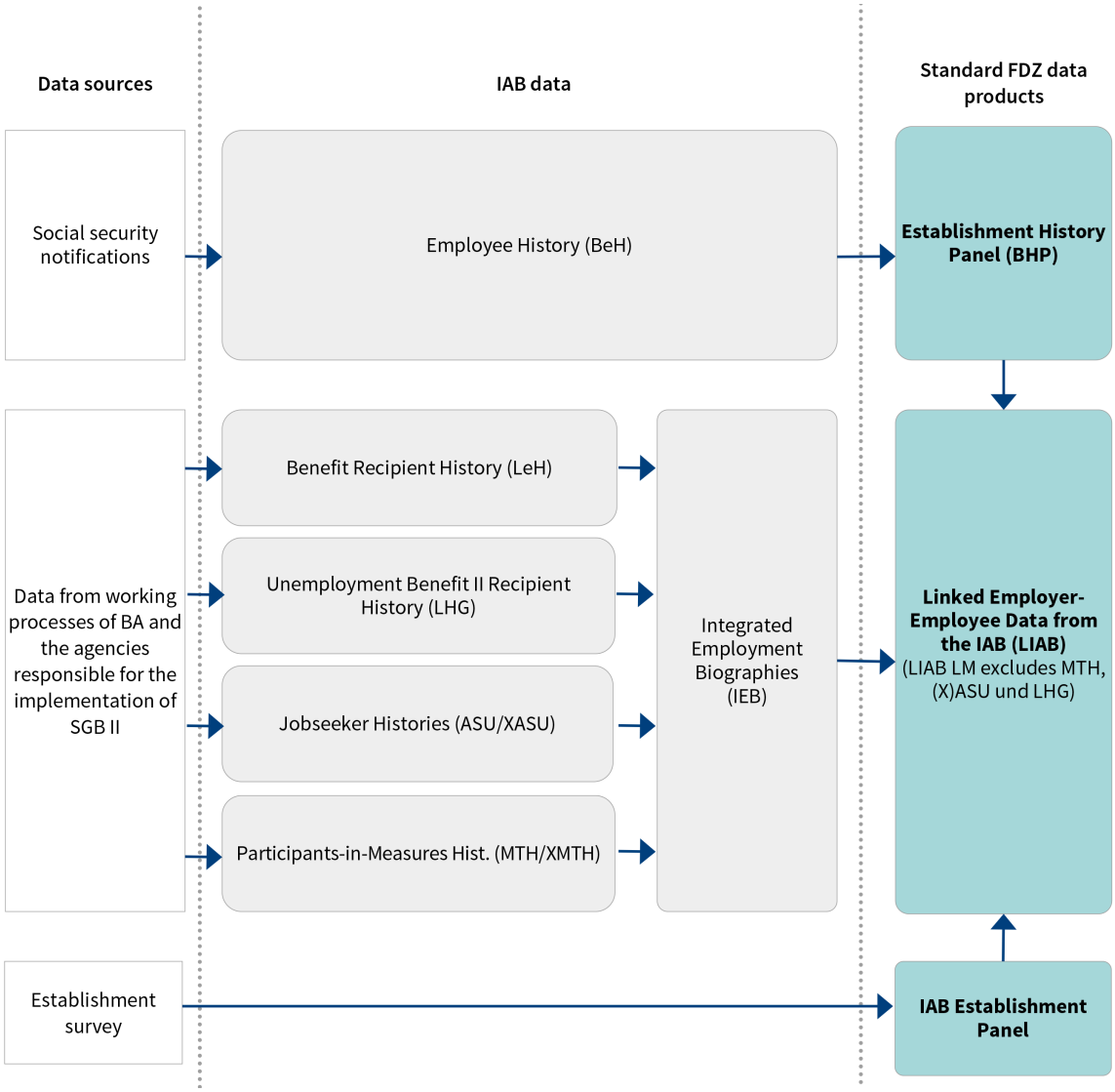



**Figure 3: Data sources of the LIAB**

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 persons and regular students[2] (see Cramer 1985) are not recorded in the BeH in principle. As the notification procedure was modified on January 1, 1999, employees in marginal part-time employment and unpaid family workers have also been recorded (not contained in the data until April 1, 1999). The data are recorded by the health insurance companies, collected and edited by the Federal Employment Agency (BA) and subsequently integrated into the History File by the IAB.

The administrative individual data are supplemented with administrative establishment data (Basis Establishment File and Extension Files). They are taken from the Establishment History Panel (Betriebs-Historik-Panel – BHP), which is also based on the BeH.

[2] Students may still appear in the BeH if, for example, they had a marginal part-time job parallel to their degree course.

When linking individual data with establishment data, it has to be taken into account that the variables in the Basis Establishment File as well as in the BHP Extension Files are aggregated as of June 30 of a year. Unlike the data on individuals, the establishment variables are therefore not spell data but are only valid on June 30 precisely each year.[3]

The Basis Establishment File is linked with the Individual File via the programme-specific commands of the software packages used for preparing and analysing the data. In Stata, for instance, the two files can be linked using the '`merge`' command in connection with the relevant paths (see Box 2).

## 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 (sphere of Social Code Book III). The benefits comprise unemployment benefit, unemployment assistance, maintenance allowance, and contributions paid by the BA to private health or care insurance while benefits are being drawn. Benefits in the context of Social Code Book II (e.g., unemployment benefit II) are not included in the data. As entitlement to benefits depends on meeting certain legal requirements, periods of unemployment in which the requirements are not met (e.g., no eligibility for unemployment assistance, or non-completion of the qualifying period for unemployment benefit) are not reported in the LeH. The earliest available data in the LeH are from January 1, 1975.

# 3 Data preparation and sampling procedure

## 3.1 Corrections and validation procedures

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 lacking information on the date of birth and gender after correction are deleted.

No further corrections (such as the addition of presumably missing notifications, strike corrections etc.) are performed.

[3] An extreme example: an employment notification exists from January 1, 2006 to May 30, 2006; the establishment goes bankrupt in June 2006. There is then no information on this establishment in the BHP for 2006.

### 3.1.1 Employee History (BeH)

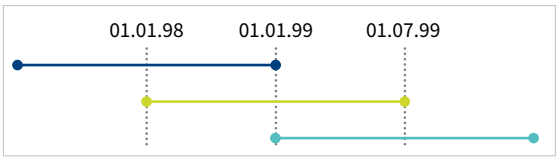
- To capture a person group that is as constant as possible over time, some person groups for which data are not available throughout the entire observation period are excluded. From the reporting year 2011 onwards, the BeH data originate from newly designed source data. As a result, a number of person groups have been introduced or reactivated as they are classified by the BA statistics as being subject to social security contributions. The person groups 101 - 107, 111 - 114, 118, 119, 120, 140, 141, 142, 143, 149, 201 and 203 - 205 are therefore contained from that time onwards as well as the two groups 109 and 209, which indicate people in marginal part-time employment. Groups that are not included are, for example, people in short-term employment, i.e. person groups 110, 202 and 210.
- Person groups 123, 124 and 127 have been newly introduced in 2011.
- For data protection reasons, the person groups 107, 111, 113, 114, 127 and 204 are combined to form the person group 'other workers' (599).
- From the reporting year 2012 onwards, apprentices were included as the new person groups 121 and 122.
- Observations with no details on earnings, the value 101 for the person group variable and the value 50 for the reason for notification (annual notification) are not incorporated into the IEB.
- Gender and date of birth are taken from the Data Warehouse (DWH) of the BA. 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 December 31, 2017.

### 3.1.2 Benefit Recipient History (LeH)

- Observations without a valid start date are excluded.
- If the end date for the receipt of unemployment assistance precedes the start date by one day and the spell was not deleted, then the end date is increased by one year.
- Between 2004 and 2006 the notification procedure from which the data originate was changed. Overlaps occurring between the old and the new procedures were corrected.
- Observations with no end date or an invalid end date are excluded, as in these cases it cannot be assumed that a benefit payment was made at all.
- The territorial allocations are corrected in the same way as in the BeH.

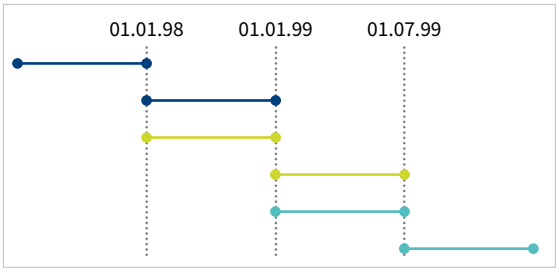
## 3.2 Episode splitting

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 (see Figure 4).




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

**Episode splitting**
Splitting of the time intervals into "non-overlapping" intervals




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

**Figure 4: Episode splitting**

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`).

It is advisable to sort entirely parallel observations generated by the splitting procedure in a consistent manner. The variables ‘observation counter per episode’ (*level2*) and ‘observation counter per episode and source’ (*level1*) that were previously contained in the LIAB models can be generated using the following Stata commands if required (see Box 3).

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

**Box 3: Example code to create additional observation counters in Stata 14**

## 3.3 Sampling procedure

The LIAB LM is sampled on the basis of the establishments surveyed in the IAB Establishment Panel. The IAB Establishment Panel is an annually conducted survey. The sample is drawn from the population of all German establishments with at least one employee liable to social security. It is stratified according to industry, firm size, and federal state.

In a first step, all so-called *panel cases* for the time period 2009 to 2016 are selected for the LIAB LM 7517. *Panel cases* comprise

- the stock of establishments in the IAB Establishment Panel in the year 2009,
- 'new' establishments in the subsequent waves,
- and establishments going out of operation during the panel period.

A prerequisite for the selection as a *panel case* is a valid interview in the following years or that the establishment goes out of operation.[4] It is important to bear in mind that not all cross-sectional cases of one wave of the IAB Establishment Panel are *panel cases* simultaneously. If one establishment was added in the course of the reserve sample, which is drawn from the stock records, or the same establishment unit wasn't surveyed as in the previous year, the corresponding case is not a *panel case* (see Fischer et al. (2008), p. 20f).

In a second step, all individuals, who were employed in one of these establishments at least one day during the time period 2008 – 2017, are identified.

In the third step, the employment biographies from the BeH and LeH for these individuals are drawn from the IEB. The employment biographies cover the time period from 1975 – 2017. Figure 5 illustrates the time dimension in the LIAB LM 7517.

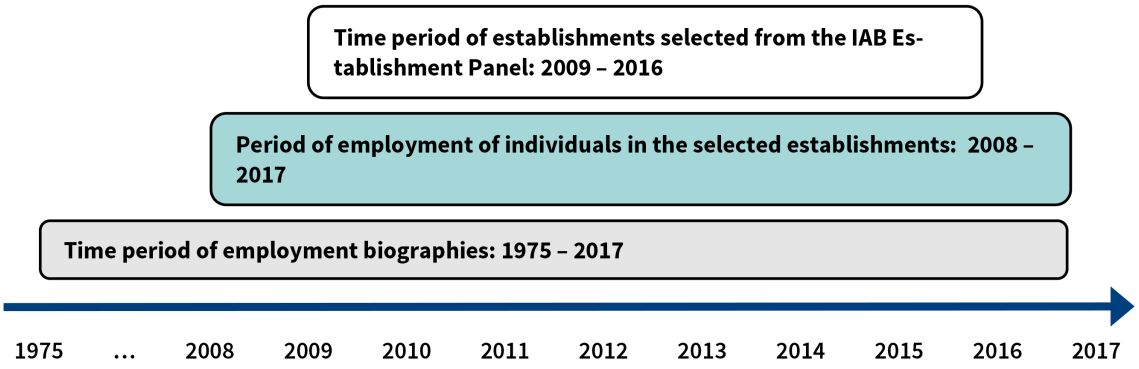



**Figure 5: Time dimension in LIAB LM 7517**

[4] See Fischer et al. (2008), p. 26f.

## 3.4 Missing values

In the LIAB, missing values are coded as follows:

| Term | Value | Description |
|---|---|---|
| No (valid) details available | .z | Values of a variable that 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 cannot be interpreted reasonably. |
| Systematically not available | .n | A variable is not available in principle for a data source (cells shaded in dark grey in the overview of Table 4) or is not available for a certain period. |

# 4 Data quality and problems

## 4.1 Entire IEB

### 4.1.1 Gaps in employment histories

The IEB contain comprehensive employment histories. However, not every type of employment is included in the administrative data. Some individuals 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 creating control groups, analysing life courses, etc.). The gaps listed below are defined as periods of time after the end of school education for which no data are included in the IEB. These gaps can be divided into

- gaps with no information available at all, and
- gaps for which information may be available from the variable reason for notification/reason for end of benefit receipt/reason for deregistration of the observation immediately preceding the gap (if a corresponding observation exists).

These gaps were identified using the variables 'Reason of cancellation/notification/termination' (*grund*) and 'Employment status' (*erwstat*) in the various sources. The list in Table 6 makes no claims to be exhaustive.

**Table 6: Biographical gaps and possible ways of identifying them**

| **Biographical gap** | **Information on gap, potentially identifiable using the details in the '*grund*' variable in the preceding observation of the source** |
|---|---|
| Civil servants, professional soldiers, judges, employees of bodies or foundations under public law | XASU |
| Self-employed persons without support | LeH, ASU |
| Students, persons in school-based further education | LeH, LHG, ASU, XASU |

| | |
|---|---|
| Persons who are ill / not able to work for more than 6 weeks (illness during unemployment, however, is represented in the ASU source under certain circumstances, see Section) | BeH, LeH, ASU |
| Persons receiving old-age pension without employment if not a member of a benefit unit | LeH, LHG, ASU |
| Individuals on maternity leave / parental leave | XASU |
| Recipients of early retirement benefits | LeH, ASU |
| Trade professionals working from home | |
| Employees working short-time | ASU |
| Persons in youth welfare facilities, in vocational training centres, approved workshops or similar facilities for disabled persons | ASU |
| Participants in programmes to support participation in working life (people in rehabilitation) | ASU |
| (Sideline) farmers | |
| Caregivers according to Section 19 SBG XI | |
| Conscripts | BeH, LeH, LHG, ASU, XASU |
| Persons in reserve duty training | BeH, LeH, LHG, ASU, XASU |
| Persons fulfilling community service | BeH, LeH, LHG, ASU, XASU |
| 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.) | BeH, LeH, ASU |
| Strikers in cases where the strike lasts more than a month | LeH |
| Social assistance recipients (prior to the introduction of SGB II in 2005), recipients of welfare payments (according to SGB II) | |
| SGB-II recipients whose providers have experienced delivery failures | |
| Recipients of compensation according to FELEG (Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit, Act on Support in Case of Termination of Farming Activities) | |

### 4.1.2 Introduction of SGB II and subsequent institutional changes

With the introduction of the SGB II on January 1, 2005, the responsibilities for the care of jobseekers were redesigned, so that tasks in the sphere of the SGB II can be taken over by different types of providers:

- Usually, the Arbeitsgemeinschaften (ARGE) between BA and the district took over the tasks according to the SGB II. These were replaced in 2011 by the joint facilities (gE) – also known as 'Job Centers'.
- In addition, it was possible until the end of 2011 for the BA and the district to perform the tasks assigned to them in their own (separate) responsibility (gT and gAw).

- Since 2005, it is also possible for authorised municipalities (zkT; also: opting municipalities) to take over the tasks. Initially, 69 administrative districts or independent towns took sole responsibility for the basic provision for jobseekers until December 31, 2010. With a constitutional amendment, the option was extended to 110 providers from 2012 onwards.

While the ARGEn/gE maintain benefit and case management via the BA procedures and transfer the recorded data to the statistics department of the BA, the municipal institutions responsible for implementing SGB II each use their own software systems. The transfer to the statistics department of the BA takes place via the XSozial standard. A division of responsibilities existed for separate providers. Any given provider was only responsible for the collection and delivery of the data falling within its area of responsibility. For this reason, there were reduced data requirements for the municipal side of the separate providers. Transmission via the XSozial standard takes place once a month at a fixed time window. In the following week, there is the possibility of repeating failed deliveries on a second reporting day. Over the years, numerous quality assurance instruments were developed. Nevertheless, version changes of the standard or the software used on site regularly increase the risk of data quality problems.

As part of the reorganisation of SGB II institutions on January 1 of the years 2011 to 2014, various ARGEn/gE and separate providers were transferred to municipal providers or converted from zkT to gE. This change of responsibility and the associated change of the software used locally also lead to some breaks in the data of the SGB-II sources. Detailed information on the resulting quality problems in the data can be found in the individual chapters on the data sources.

## 4.2 Employee History (BeH)

- Information on vocational training, the occupation/activity performed and the occupational status is transmitted by means of notifications made by the employer in accordance with the Data Collection and Transmission Regulation (DEÜV) (see Section 2.1) using a so-called occupation code. The new occupation code 2010 was adopted for notifications with an end date later than November 30, 2011 (for further details, see Bertat et al., 2013). The decision to switch to the new occupation code was made by the central organisations of the social security agencies as a number of facts could no longer be recorded in a way that was up-to-date and realistic using the occupation code 2003. As the notifications made by employers in accordance with DEÜV only enter the Employee History (BeH), the change of the occupation code only affects observations from this source. The measurement of the following characteristics previously reported using the occupation code 2003 is affected by this change: working hours, occupation, occupational status and school and vocational qualification levels. In addition, since the switch to the new occupation code, details are also available about whether an employment relationship is fixed-term and whether a person is employed by a temporary work agency to be hired out to other firms. The most important consequence is the switch to a new occupational classification. Instead of the previous Classification of Occupations 1988 (Klassifikation der Berufe 1988 (KldB 1988)), the more highly differentiated KldB 2010 is reported with the new occupation code.[5]

[5] See Paulus/Matthes, 2013, for details regarding the Classification of Occupations 2010.

- In contrast to LIAB LM 9314, LIAB LM 7517 contains the variable 'Occupational status and working hours' (*stib*) again. However, the variable is only filled for reports that were submitted before the introduction of the new occupation code, and the FDZ does not extrapolate or impute for later reports. The categories of the variable *stib* that can be consistently observed over the whole observation period (occupational status as a trainee, distinction between part-time and full-time) are to be reproduced in the variables *erwstat* and *teilzeit*.
- The introduction of the new occupation code in 2011 led to a number of problems. For example, during the transition period granted to employers in the social security notification procedure,[6] there was a temporary increase in the number of missing details. Analyses of the BA statistics (Bertat et al., 2013, p. 10) show that in 20 to 30 percent of cases no information was available in the new or converted variables 'Occupation/activity performed', 'Working time' and 'Vocational education and training' after the switch. This situation began to improve significantly in the first half of 2013. In order to improve the quality of the 'Working time' variable in the transition period, Ludsteck/Thomsen (2016) developed an imputation procedure to replace the missing values by imputed values. The imputed data are included in the LIAB LM 7517. No imputation is performed regarding the gaps in the other variables.
- 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 large number of spells for 1991 have missing values for several variables (such as 'Vocational training', 'Employment status', and 'Daily wage').
- 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 April 1, 1999 onwards.
- Especially in 1999, observations of part-time employment increase 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 correctly.
- 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 (until the annual notification 2012) or mid-February (from the annual notification 2013 onwards) 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-months file for 2013 can be created in July 2015 at the earliest). Notifications submitted more than three years late are not taken into account at the IAB, which means that a 36-months file shows a 100 percent degree of completeness by definition.
- In the version of the IEB on which the LIAB LM 7517 is based, the year 2014 is the last year with a degree of completeness of BeH observations of 100 percent.[7] For the years 2015 and 2016,

[6] The test programs used in the notification procedure permitted missing details in the occupation code 2010 until the end of May 2012.

[7] Due to a redesign of the data basis of the BeH, at the time of the preparation of the BeH as an exception only the 30-months files were available instead of the 36-months files. However, analyses with earlier data versions have shown that, as a rule, the

the 18-months files were used, and the observations for 2017 originate from a 6-months file. It can therefore be assumed that employment notifications for 2015 and 2016 are slightly underreported in the LIAB, and that those for 2017 are underreported to a slightly higher degree. However, this should not reduce the ability to analyse the data at 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 leads to an increase in the average daily wage. In particular, the proportion of wages and salaries above the upper earnings limit increased considerably from that year onwards (see Bender et al. 1996).
- For the years 1992 until 2000, noticeable decreases and increases in the number of notifications were observed. Decreases can be observed especially for the following 10 districts: Braunschweig (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 one or more establishments in these regions.
- Considerable decreases were also observed for the districts Salzgitter (03102) and Hoyerswerda (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 at one or more establishments in these regions.
- In the years 1996 to 1998, the values 841-844 (doctors and pharmacies) within the 'Occupation – activity performed' (*beruf*) variable are very rare compared to the neighbouring years. The reasons for this are not known.
- In the years 1975 and 1977, there were so far considerably too many employees with a place of work municipality of Lahn (district of Emsland, Lower Saxony). Instead of the expected double-digit figure, there were up to 90,000 employee registrations per year with this place of work in the population of the BeH. The reason for this is a historical misclassification of employment reports from the city of Lahn. This was a merger of the Hessian municipalities Wetzlar, Gießen, Heuchelheim, Wettenberg and Lahnau (districts of Gießen and Lahn-Dill-Kreis) which was dissolved after a short time. On the basis of these findings and assuming that there were no real establishment relocations between these regions, the following adjustment rule was implemented at the level of the establishment number: As soon as the establishment location changes from the municipality of Lahn to one of the listed Hessian municipalities in the years 1975 to 1978, the former specification of the municipality of Lahn is overwritten with the later correct specification. This rule significantly reduces the overhang and the municipality of Lahn

18-months file already has a degree of completeness of around 99 percent. This means that in the next 18 months there will only be extremely few follow-up or correction notifications.

in Lower Saxony then only has less than 2,000 employees in the population of the BeH in the critical years.

## 4.3 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 contained. Thus, in the present data product, too, it can be assumed that information on benefit receipt in that period is not available in full.
- Due to an internal change of systems, there is a break in the recording of periods of exclusion from benefits and of benefit suspension in 2004. Until July 1, 2004 periods of exclusion from benefits and of benefit suspension can only be identified via the 'Reason for end of benefit receipt' in the preceding LeH observation. After this date, a separate observation is available with the daily benefit rate = 0 for periods of benefit exclusion and suspension.

# 5 Description of variables

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

## 5.1 Identifiers

### 5.1.1 Individual ID (persnr)

| | |
|---|---|
| Variable label | Individual ID |
| Variable name | persnr |
| Category | identifiers |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | The individual ID 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.<br>As there is no uniform individual identifier in the different data sources, 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.<br>The formation of the individual identifier which spans all data sources is based on a heuristic developed by the BA. |

### 5.1.2 Establishment ID (betnr)

| | |
|---|---|
| Variable label | Establishment ID |
| Variable name | betnr |
| Category | identifiers |

| | |
|---|---|
| Origin | BeH |
| Data type | numerical |
| Detailed description | The establishment ID indicates which observations belong to the same establishment. It is based on the establishment number allocated by the BA, which was replaced by an artificial number (further information on the allocation of establishment numbers by the BA can be found in Bender et al. (1996: p. 15 et seq. and pp. 27-30) as well as directly on the website of the establishment number service of the BA at https://www.arbeitsagentur.de/betriebsnummern-service/alles-wichtige). The establishment number and year specification can be used to merge individual and establishment information.<br><br>For the establishment number, the following should be observed in general:<br>If the company has only one office, or if the company has only one office in one municipality, this office is the establishment and is given an establishment number.<br>If the company has **several** branch offices in **one** municipality, these establishment premises / workplaces must be merged into a single establishment under one establishment number, if they belong to the same economic class. If they do **not** belong to the same economic class, each branch office is regarded as a separate establishment and is given its own establishment number.<br>If the company has **several** branch offices in **several** municipalities, each of these branch offices is an establishment and is given its own establishment number.<br><br>In this context, the following definitions with regards to the allocation of establishment numbers as part of the notification procedure for social security must be observed:<br>An **establishment** is a regionally and economically delimited unit in which employees work and which is allocated an establishment number according to the above-mentioned principles.<br>A **workplace** is a unit in which employees work and which is not allocated an establishment number according to the above-mentioned principles.<br>A **company** as a term combines establishment premises and workplaces belonging to the same employer.<br>An **employer** is any natural person or legal entity that employs at least one employee subject to social security contributions or in marginal part-time employment.<br>Establishment and establishment premises are synonyms; branch office is a synonym for subsidiary, district office, out-sourced office, workplace etc. if it is not an establishment. |
| Notes on quality | The establishment ID is only missing in a very small number of cases. These observations are notifications for the person group '205' (earnings notifications for casual workers). As establishment variables (place of work, economic activity, establishment size etc.) are merged via the establishment ID, they are missing in these observations. |

### 5.1.3 Establishment-ID acc. to Survey (idnum)

| | |
|---|---|
| Variable label | Establishment-ID acc. to Survey |
| Variable name | idnum |
| Category | identifiers |
| Origin | IAB Establishment Panel |
| Data type | numerical |

| | |
|---|---|
| Detailed description | The Establishment ID acc. to Survey is adopted from the IAB Establishment Panel. The variable is only filled in the Individual File if the person is employed in an establishment which is in the IAB Establishment Panel on the reference date with a valid interview. |

## 5.2 Generated technical variables

### 5.2.1 Counter per person (spell)

| | |
|---|---|
| Variable label | counter per person |
| Variable name | spell |
| Category | generated technical variables |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | The observation counter per person counts a person's observations in an ascending order. The variable is generated during the episode splitting procedure and refers to the split observations. Using the 'observation counter per person' variable, it is straightforward to restore the original sorting order. The observations are sorted by the start date of the split episode first and then by the data source second. Within employment notifications, persons subject to social insurance contributions are sorted before marginal employment notifications and higher daily wages before lower ones. One exception constitute one-time payments, which are sorted last. |

### 5.2.2 Source of spell (quelle)

| | |
|---|---|
| Variable label | Source of spell |
| Variable name | quelle |
| Category | generated technical variables |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | The variable indicates the data source (see Section 2). |

### 5.2.3 Year (jahr)

| | |
|---|---|
| Variable label | Year |
| Variable name | jahr |
| Category | generated technical variables |
| Origin | BeH |
| Data type | numerical |
| Detailed description | This variable indicates the calendar year at the beginning of each episode (i.e. after the episode splitting, see Section 3.2).<br>This variable can be used together with the establishment number to link the Individual File and the establishment files. See Box 2 on page 10 for an example code with the '`merge`' command in Stata 14. |

### 5.2.4 Status of establishment number (betr_st)

| | |
|---|---|
| Variable label | Status of establishment number |
| Variable name | betr_st |
| Category | generated technical variables |

| | |
|---|---|
| Origin | BeH |
| Data type | numerical |
| Detailed description | For the current episode in the individual data, the variable indicates if there is information from the IAB Establishment Panel for this episode. The following characteristics exist:<br><br>1. Employed June 30 of year with valid interview, establishment selected for LIAB model<br>2. Not employed June 30 of year with valid interview, establishment selected for LIAB model<br>3. Employed in establishment of Establishment Panel, establishment not selected for LIAB model<br>4. Employed in establishment not part of Establishment Panel<br><br>The values 1-3 indicate employment in establishments which have taken part in the IAB Establishment Panel survey at least once, whereas value 4 marks employment in establishments which have never taken part in the IAB Establishment Panel.<br>Values 1 and 2 indicate employment in establishments which are included in the IAB Establishment Panel and which have been selected for the respective LIAB model. Value 3 defines employment in establishments which have not been selected for the respective LIAB model.<br>The difference between values 1 and 2 depends on whether the episode includes June 30 of the respective year for a panel establishment that has been interviewed.<br>The value 3 does not appear in a cross-sectional model of the LIAB by construction. |

## 5.3 Period of validity

### 5.3.1 Original start date (begorig)

| | |
|---|---|
| Variable label | Original start date |
| Variable name | begorig |
| Category | period of validity |
| Origin | BeH, LeH |
| Data type | date |
| Detailed description | The original start date of the observation corresponds to the original start date of the notification. This date can differ from the start date of the episodes (*begepi*) (see also the comments on episode splitting in Section 3.2).<br>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). A continuous employment relationship may therefore be distributed across several notifications. |

### 5.3.2 Original end date (endorig)

| | |
|---|---|
| Variable label | Original end date |
| Variable name | endorig |
| Category | period of validity |
| Origin | BeH, LeH |
| Data type | date |

| | |
|---|---|
| Detailed description | The original end date of the observation corresponds to the original end date of the notification. This date 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 Section 3.2).<br>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). A continuous employment relationship may therefore be distributed across several notifications. |

### 5.3.3 Episode start date (begepi)

| | |
|---|---|
| Variable label | Episode start date |
| Variable name | begepi |
| Category | generated period of validity |
| Origin | BeH, LeH |
| Data type | date |
| Detailed description | The start date of the split episode is always equal to or greater than the start date of the original observation (see also the comments on episode splitting in Section 3.2). |

### 5.3.4 Episode end date (endepi)

| | |
|---|---|
| Variable label | Episode end date |
| Variable name | endepi |
| Category | generated period of validity |
| Origin | BeH, LeH |
| Data type | date |
| Detailed description | The end date of the split episode is always equal to or smaller than the end date of the original observation (see also the comments on episode splitting in Section 3.2) |

## 5.4 Personal information

### 5.4.1 Gender (frau)

| | |
|---|---|
| Variable label | Gender |
| Variable name | frau |
| Category | personal variable |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | Gender dummy (0 - man, 1 - woman). The gender information is constant within one individual account. |

### 5.4.2 Year of birth (gebjahr)

| | |
|---|---|
| Variable label | Year of birth |
| Variable name | gebjahr |
| Category | personal variables |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | The year of birth is constant within one individual account. |

| | |
|---|---|
| Notes on quality | In the original data, it may happen that the date 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. |

### 5.4.3 Month of birth (gebmon)

| | |
|---|---|
| Variable label | Month of birth |
| Variable name | gebmon |
| Category | personal variables |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | The month of birth is constant within one individual account.<br>One can use the variables 'Year of birth' (*gebjahr*) and 'Month of birth' (*gebmon*) to generate a variable in the date format JJJJmM (e.g., 1984m6) with the following syntax in Stata:<br>`gen int gebdat = ym(gebjahr, gebmon)`<br>`format gebdat %tm` |
| Notes on quality | In the original data, it may happen that the date 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. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable 'Year of birth' (*gebjahr*) is provided. |

### 5.4.4 Nationality (nation)

| | |
|---|---|
| Variable label | Nationality |
| Variable name | nation |
| Category | personal variables |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | The variable contains the nation codes used by the Federal Statistical Office (Statistisches Bundesamt, 2019). |
| Notes on quality | The variable is not filled well in the LeH before 1983. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable 'Nationality, grouped' (*nation_gr*) is provided. |

### 5.4.5 Nationality, grouped (nation_gr)

| | |
|---|---|
| Variable label | Nationality, grouped |
| Variable name | nation_gr |
| Category | personal variables |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | The variable contains a grouped version of the nation codes used by the Federal Statistical Office (Statistisches Bundesamt, 2019). |
| Notes on quality | The variable is not filled well in the LeH before 1983. |

### 5.4.6 Marital status (famst)

| | |
|---|---|
| Variable label | Marital status |
| Variable name | famst |
| Category | personal variables |
| Origin | LeH |
| Data type | numerical |
| Detailed description | This variable describes the marital status. In the LeH, the variable has only two values (0 - not married, 1 - married). |
| Notes on quality | The quality of the information originating from the LeH is classified as poor until 2005, as it is not quite clear how the marital status was maintained. The quality is also questionable since the beginning of 2014. Between 2013 and 2017, for example, the proportion of married people drops from 37% to 14%. |

### 5.4.7 Number of children (kind)

| | |
|---|---|
| Variable label | Number of children |
| Variable name | kind |
| Category | personal variables |
| Origin | LeH |
| Data type | numerical |
| Detailed description | In the LeH, the variable indicates the number of children aged under 16 living in the household at the time when the application was made. It only distinguishes between whether the number of children is zero or at least one. The following values are therefore available:<br>0 no children<br>100 one or more children |
| Notes on quality | In the LeH, the variable is not updated when there are changes in the type of benefit or the approval of benefits, but only when a new case of benefit receipt occurs after a period of employment. This strongly restricts the quality of the data. |

### 5.4.8 Vocational training (ausbildung)

| | |
|---|---|
| Variable label | Vocational training |
| Variable name | ausbildung |
| Category | personal variables |
| Origin | BeH |
| Data type | numerical |
| Detailed description | For observations obtained from the BeH, the variable contains the vocational education reported by the employers as part of the employment notification procedure. The following values exist:<br><br>1 Without vocational training<br>2 In-company voc. training/traineeship/external voc. training<br>11 University of applied sciences without further specifications<br>12 University without further specifications<br><br>In notifications that rely on the new occupation code (see Section 4.2) it is no longer possible to identify graduates of universities of applied sciences clearly, as the new |

| | |
|---|---|
| | occupation code no longer has a separate category for this vocational qualification. They are assigned to category 12. |
| Notes on quality | "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 obtained 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 carried forward" (own translation of Meinken / Koch 2004, p. 63).<br>In the BeH, the share of missing values increases almost continuously over time. Due to the introduction of the new occupation code in 2011, the share even temporarily strongly increased to around 51%. Since 2014, however, the proportion of missing values in the BeH has levelled off at around 40%.<br>Missing values occur particularly frequently in the following groups: marginal part-time employees, part-time workers, foreign employees and workers from Eastern German. The reason for this is that the variable is not of particular importance as regards social security contributions (see Meinken/Koch, 2004, p. 63).<br>For the variable 'Vocational training (imputed)' (*ausbildung_imp*, see Section 5.4.9), a method was applied to correct missing values or inconsistent changes of the training variable. However, this variable is only filled in the source BeH and has different categories than the variable *ausbildung*. |

### 5.4.9 Vocational training (imputed) (ausbildung_imp)

| | |
|---|---|
| Variable label | Vocational training (imputed) |
| Variable name | ausbildung_imp |
| Category | personal variables |
| Origin | BeH |
| Data type | numerical |
| Detailed description | The variable 'Vocational training (imputed)' (*ausbildung_imp*) is a supplement to the variable 'Vocational training' (*ausbildung*) and contains additional and harmonised information on the vocational training of employees for BeH spells. The variable thus offers a solution to the problems concerning the variable *ausbildung* described in Section 5.4.8. The imputation procedure is described in Thomsen et al (2018).<br>As the variable only uses the training information from BeH notification and because the educational categories of the old and the new occupation codes had to be harmonised for the variable *ausbildung*, the variable *ausbildung_imp* has other categories than the variable *ausbildung*. |

### 5.4.10 School leaving qualification (schule)

| | |
|---|---|
| Variable label | School leaving qualification |
| Variable name | schule |
| Category | personal variables |
| Origin | BeH |
| Data type | numerical |
| Detailed description | With the switch to the new occupation code (see Section 4.2), the possible values of the variable change. In the BeH, the values from the old occupation code are: |

| | |
|---|---|
| | 5 Grade-/lower school certificate, intermediate school or equivalent qualification<br>8 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college or upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling<br>9 Upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling<br><br>With the new occupation code the values are:<br><br>1 No school leaving certificate<br>4 Lower secondary school certificate/ grade school certificate<br>6 Intermediate school leaving certificate<br>8 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college or upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling |
| Notes on quality | The degree of completeness in the BeH has been decreasing continuously over time and seems to have levelled off at under two thirds in recent years. |

## 5.5 Information on employment and benefit receipt

### 5.5.1 Daily wage/daily benefit (tentgelt)

| | |
|---|---|
| Variable label | Daily wage/daily benefit |
| Variable name | tentgelt |
| Category | information on employment and benefit receipt |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | **1) BeH**<br>In BeH observations, this variable shows the employee's average gross daily wage. It is calculated from the fixed-period earnings reported by the employer and the duration of the (unsplit) original notification period in calendar days. The daily wage is shown in Euros.<br>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. Earnings exceeding the upper earnings limit for statutory pension insurance are only reported up to this limit. There are two upper earnings limits in the statutory pension insurance scheme. The earnings limit of the miners' pension insurance is generally higher than the earnings limit of the pension insurance for wage and salary earners. However, it is not possible to differentiate between these two insurance providers in the data.<br>Since the inclusion of marginal part-time employees in the employment notification procedure on April 1, 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. In this case, it is irrelevant whether the upper earnings limit in the statutory pension insurance which is decisive for the notification period is exceeded as a result of this addition. However, such |

| | |
|---|---|
| | earnings notifications could also be due to incorrect details in the employment period. The earnings in-formation, however, may be considered less error-prone due to its insurance relevance.<br>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 https://fdz.iab.de/en.aspx.<br>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.<br>From 2013 onwards, the number of notifications with a reason for deregistration of 54 (notification of a one-off wage) increases sharply (see Section 5.5.12). It is likely that special payments which were reported with the annual declarations before 2013 are now reported separately. It is therefore advisable to add these variable one-time payments to the corresponding wages for simultaneous employment episodes within the same establishment when analysing wages over time.<br>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.<br><br>**2) LeH**<br>For LeH observations, the variable shows the daily benefit rate, converted into Euros in each case. It must be taken into account that for observations with an original end date prior to January 1, 1998 the daily benefit rate applies to working days (i.e., including Saturdays but excluding Sundays and public holidays), while for observations with an original end date from January 1, 1998 onwards it applies to calendar days.<br>Since January 1, 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 observations that reflect a period of benefit suspension, the entitlement is the same as before the start of the benefit suspension period. |

### 5.5.2 Occupation - current/most recent (KldB 1988) (beruf)

| | |
|---|---|
| Variable label | Occupation – current/most recent (KldB 1988) |
| Variable name | beruf |
| Category | information on employment and benefit receipt |
| Origin | BeH |
| Data type | numerical |
| Detailed description | In the BeH, the occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. 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).<br>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. |

| | |
|---|---|
| | Employment notifications with an end date later than November 30, 2011 are reported using the new occupation code 2010 (KldB2010) (see Section 4.2). These values are transcoded to the KldB1988 via a priority switch. This results in inaccuracies. |
| Notes on quality | In the BeH, there is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. A similar accumulation of missing values occurs in 1991 due to reunification. |

### 5.5.3 Occupational group - current/most recent (KldB 2010), 3-digit (beruf2010_3)

| | |
|---|---|
| Variable label | Occupational group - current/most recent (KldB 2010), 3-digit |
| Variable name | beruf2010_3 |
| Category | information on employment and benefit receipt |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | **1) BeH**<br>The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes apply for one employee, the employer is required to select the job title that best defines the main activity performed (see Bundesagentur für Arbeit, 2005, p. V).<br>For this the employer encodes the employee's job in accordance with the 'Classification of Occupations 2010' (Klassifikation der Berufe 2010, KldB2010, Bundesagentur für Arbeit, 2011). The occupational class consists of a 5-digit code and comprises about 1,300 values. The less detailed occupational group is recorded by the first three digits of the code. The skill level required for a job, which is recorded in the fifth digit of the codes in the KldB2010, is made available separately in the variable 'Level of requirement' (*niveau*).<br>Employment notifications with an end date earlier than November 30, 2011 are reported using the old occupation code 1988 (KldB 1988) (see Section 4.2). These values are recoded to the KldB2010 by transferring the key area. As the new occupation code is considerably more detailed than the old one, this results in substantial inaccuracies. This must be taken into account when analysing the data.<br><br>**2) LeH**<br>The variable contains the occupation of the last job. See 1) with regard to the occupation code. |
| Notes on quality | **1) BeH**<br>There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. A similar accumulation of missing values occurs in 1991 due to reunification.<br><br>**2) LeH**<br>This variable was previously not filled in the LeH because the occupational data in that source was of poor quality. By now, the occupation in the LeH is adopted from the jobseeker history, leading to an increase in quality. |

### 5.5.4 Occupational sub-group - current/most recent (KldB 2010), 4-digit (beruf2010_4)

| | |
|---|---|
| Variable label | Occupational sub-group - current/most recent (KldB 2010), 4-digit |
| Variable name | beruf2010_4 |
| Category | information on employment and benefit receipt |
| Origin | BeH, LeH |

| | |
|---|---|
| Data type | numerical |
| Detailed description | **1) BeH**<br>The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes apply for one employee, the employer is required to select the job title that best defines the main activity performed (see Bundesagentur für Arbeit, 2005, p. V).<br>For this the employer encodes the employee's job in accordance with the 'Classification of Occupations 2010' (Klassifikation der Berufe 2010, KldB2010, Bundesagentur für Arbeit, 2011). The occupational class consists of a 5-digit code and comprises about 1,300 values. The less detailed occupational sub-group is recorded by the first four digits of the code. The skill level required for a job, which is recorded in the fifth digit of the codes in the KldB2010, is made available separately in the variable 'Level of requirement' (*niveau*).<br>Employment notifications with an end date earlier than November 30, 2011 are reported using the old occupation code 1988 (KldB 1988) (see Section 4.2). These values are recoded to the KldB2010 by transferring the key area. As the new occupation code is considerably more detailed than the old one, this results in substantial inaccuracies. This must be taken into account when analysing the data.<br><br>**2) LeH**<br>The variable contains the occupation of the last job. See 1) with regard to the occupation code. |
| Notes on quality | **1) BeH**<br>There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. A similar accumulation of missing values occurs in 1991 due to the reunification.<br><br>**2) LeH**<br>This variable was previously not filled in the LeH because the occupational data in that source was of poor quality. By now, the occupation in the LeH is adopted from the jobseeker history, leading to an increase in quality. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable *beruf2010_3* is provided. |

### 5.5.5 Level of requirement - current/most recent job (KldB 2010) (niveau)

| | |
|---|---|
| Variable label | Level of requirement - current/most recent (KldB 2010) |
| Variable name | niveau |
| Category | information on employment and benefit receipt |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailbeschreibung | **1) BeH**<br>The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes apply for one employee, the employer is required to select the job title that best defines the main activity performed (see Bundesagentur für Arbeit, 2005, p. V).<br>For this the employer encodes the employee's job in accordance with the 'Classification of Occupations 2010' (Klassifikation der Berufe 2010, KldB2010, Bundesagentur für Arbeit, 2011). The occupational class consists of a 5-digit code and comprises |

| | |
|---|---|
| | about 1,300 values. The less detailed occupational sub-group is recorded by the first four digits of the code. The skill level required for a job, which is recorded in the fifth digit of the codes in the KldB2010, is made available separately in the variable 'Level of requirement' (*niveau*).<br>Employment notifications with an end date earlier than November 30, 2011 are reported using the old occupation code 1988 (KldB 1988) (see Section 4.2). These values are recoded to the KldB2010 by transferring the key area. As the new occupation code is considerably more detailed than the old one, this results in substantial inaccuracies. This must be taken into account when analysing the data.<br><br>**2) LeH**<br>The variable contains the occupation of the last job. See 1) with regard to the occupation code. |
| Notes on quality | **1) BeH**<br>There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. A similar accumulation of missing values occurs in 1991 due to the reunification.<br><br>**2) LeH**<br>This variable was previously not filled in the LeH, as the occupational data in that source was of poor quality. By now, the occupation in the LeH is adopted from the jobseeker history, leading to an increase in quality. |

### 5.5.6 Part-time (teilzeit)

| | |
|---|---|
| Variable label | Part-time |
| Variable name | teilzeit |
| Category | information on employment and benefit receipt |
| Origin | BeH |
| Data type | numerical |
| Detailed description | The variable 'Part-time' (*teilzeit*) distinguishes 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. For part-time employees, the variable only records whether their working hours exceed a certain limit or not. Until 1978, this limit was 20 hours of work per week, between 1979 and 1987 it was 15 hours per week and since 1988 it has been 18 hours per week. |
| Notes on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. In order to reduce this problem, the working hours were imputed at the IAB for the period in question. Further information about the procedure can be found in Ludsteck/Thomsen (2016). A similar accumulation of missing values occurs in 1991 due to the reunification. No imputation is performed here. |

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

| | |
|---|---|
| Variable label | Occupational status and working hours |
| Variable name | stib |
| Category | information on employment and benefit receipt |
| Origin | BeH |
| Data type | numerical |
| Detailed description | The employee's occupational status during the notification period is reported by the employer as part of the 'employment details'. The variable 'Occupational status and |

| | |
|---|---|
| | working hours' (*stib*) distinguishes between full-time and part-time employees. The decisive factor for this differentiation is the ratio between the contracted hours and the usual working hours in the establishment. For part-time employees, the variable only indicates whether their working hours exceed a certain threshold. Until 1978, this threshold was 20 hours of work per week, between 1979 and 1987 it was 15 hours per week and since 1988 it is 18 hours per week.<br>The variable only provides information regarding the occupational status for full-time employees, distinguishing among other things between blue-collar and white-collar employees in full-time employment and apprentices. The distinction between (full-time) blue-collar and white-collar employees is solely based on the type of compulsory pension insurance (Federal Social Insurance Office for Salaried Employees – BfA – for white-collar workers, and State Social Insurance Office – LVA – for blue-collar workers). Master craftsmen and foremen are only included in a separate category if they are compulsorily insured in the workers' pension insurance. The assignment of a master craftsman or foreman to the blue-collar or white-collar employees can only be made via the respective pension provider. The 'Employees in vocational training' category covers not only apprentices, volunteers and interns but also semi-skilled trainees, students at colleges for health occupations and participants in subsidised further vocational training, retraining and induction training.<br>If more than one code is eligible for an employee, the employer is required to classify the job according to the activity which is predominantly performed. If this cannot be determined clearly, the code of the higher occupational status is to be entered (see BA 2005, p. VI).<br>Owing to the introduction of the new occupation code (see Section 4.2), however, this distinction is no longer possible. The variable *stib* is therefore only filled for notifications which date back to before the introduction of the new occupation code. |
| Notes on quality | There is a considerable number of missing values in 1991 due to the German reunification. |

### 5.5.8 Employment status (erwstat)

| | |
|---|---|
| Variable label | Employment status |
| Variable name | erwstat |
| Category | information on employment and benefit receipt |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | This variable takes on different values with different meanings for each data source.<br><br>**1) BeH**<br>For BeH observations, the variable 'Employment status' (*erwstat*) corresponds to the person group recorded in the new notification procedure (DEÜV) from January 1, 1999 onwards. It indicates contribution- or benefit-related particularities of the employment relationship.<br>If multiple codes 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 distinctive characteristics, which are recorded under code number 101. Accordingly, it is possible that these employment relationships are slightly overestimated.<br>The notification procedure stipulates that changes in the employment status - e.g., when an apprentice is taken on by his/her training company after completing his/her vocational training - must be indicated by a new notification.<br>The person group 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 variables 'Vocational education and training', 'Occupational status and working hours' and 'Occupation' as well as other information. In many cases, however, conclusive allocations are not possible.<br>Since April 1, 1999, employees in marginal part-time employment have also been recorded in the DEÜV notification procedure. This person group can be distinguished via the values 109 and 209. For employees in marginal part-time employment, no data prior to the introduction of the notification obligation in 1999 could be collected.<br><br>**2) LeH**<br>For LeH observations, the variable 'Employment status' (*erwstat*) contains the grouped benefit type. Thus, it is possible to distinguish 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. |

### 5.5.9 Transition zone (gleitz)

| | |
|---|---|
| Variable label | Transition zone |
| Variable name | gleitz |
| Category | information on employment and benefit receipt |
| Origin | BeH |
| Data type | numerical |
| Detailed description | This variable is only available from 2003 onwards and only for BeH observations. It indicates whether the employment notification relates to employment in the low-wage sector, within the so-called transition zone. Jobs in the transition zone have a gross monthly wage of € 400.01 to € 800.00 (so-called midi jobs) for which the employee only has to pay a reduced overall social security contribution. As employees with earnings in the transition zone can voluntarily pay the 'regular' social security contribution, not all employees with corresponding earnings are automatically classified as being in the transition zone. The corresponding legislation has been in force since April 1, 2003. |

### 5.5.10 Temporary agency work (leih)

| | |
|---|---|
| Variable label | Temporary agency work |
| Variable name | leih |
| Category | information on employment and benefit receipt |
| Origin | BeH |
| Data type | numerical |
| Detailed description | The variable reports whether the person's employment is a temporary job via an employment agency. The variable is derived from the occupation code 2010 and is only available for notifications with an end date later than November 30, 2011. |
| Notes on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. By 2012, the share of missing values is down to about 3%. |

### 5.5.11 Fixed-term contract (befrist)

| | |
|---|---|
| Variable label | Fixed-term contract |
| Variable name | befrist |

| | |
|---|---|
| Category | information on employment and benefit receipt |
| Origin | BeH |
| Data type | numerical |
| Detailed description | The variable reports whether the person's employment relationship is fixed-term or permanent. The variable is derived from the occupation code 2010 and is only available for notifications with an end date later than November 30, 2011. |
| Notes on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. By 2012, the share of missing values is down to about 3%. |

### 5.5.12 Reason of cancellation/notification/termination (grund)

| | |
|---|---|
| Variable label | Reason of cancellation/notification/termination |
| Variable name | grund |
| Category | information on employment and benefit receipt |
| Origin | BeH, LeH |
| Data type | numerical |
| Detailed description | **1) BeH**<br>In BeH observations, the 'Reason of notification' variable indicates the reason why the employer submitted the employment notification in question to the social security agencies. However, not all of the possible reasons for submitting a notification that may occur in the context of the notification procedure are available in the IEB. For instance, the IEB only includes notifications that have information on earnings (i.e., annual, employment interruption and end of employment notifications), while initial registrations are not contained 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.<br>The reasons for submitting employment notifications are encoded according to the regulations of the notification procedure, which has been in effect since January 1, 1999 (in accordance with DEÜV).<br><br>**2) LeH**<br>In the LeH, this variable specifies the reason for the end of the receipt of unemployment benefits, unemployment assistance, or maintenance allowance. There is no information in the LeH on the reasons for the start of the benefit receipt because the LeH is filled with the notifications from the employment agencies to the health insurance about completed benefit receipt durations. |
| Notes on quality | In the BeH, from 2013 onwards the number of notifications with a reason for deregistration of 54 (notification of a one-off payment) increases sharply. As long as an employment relationship exists, special payments that are paid out by March of the following year can be included into the usual notifications (mostly annual notifications) for the previous year. A separate notification with a reason for deregistration of 54 is then not required. Until 2012, the annual notifications could be submitted until mid-April; since 2013, they must now be submitted by mid-February at the latest. Special payments made in February and March must now be reported separately. |

### 5.5.13 Residual claim/planned duration (restanspruch)

| | |
|---|---|
| Variable label | Residual claim/planned duration |
| Variable name | restanspruch |
| Category | information on employment and benefit receipt |

| | |
|---|---|
| Origin | LeH |
| Data type | numerical |
| Detailed description | In the LeH, the variable contains the residual entitlement to unemployment benefit that remains after the end of the current benefit receipt period. If the period of benefit receipt ends before the maximum duration of entitlement has been reached (e.g., due to taking up employment again), a residual entitlement remains which is equivalent to the duration of benefit entitlement that was not used up. If new entitlement is acquired within five years, the duration of the residual entitlement is added to the new duration of entitlement. However, the maximum duration of entitlement for the client's age is the upper limit. If no new entitlement is acquired, the residual entitlement can be used for benefits within four years on application.<br>If the end date of the benefit receipt is before January 1, 1998, the remaining entitlement is reported in working days, after this date it is reported in calendar days.<br>This information does not refer to the start date of the episode, but to the start of the original time period. |
| Notes on quality | In the LeH, over the course of a benefit receipt biography a systematic development of the remaining entitlement is to be expected. This means that the remaining entitlement at the end of a benefit period minus the duration of the subsequent period should give the remaining entitlement after the subsequent period as long as no claims have arisen or expired in the meantime. However, there may be unexpected increases or decreases in the remaining entitlement.<br>These are presumably mainly caused by corrections during the processing of benefit cases. Such corrections occur, for example, if a claim was not correctly determined at the start of the benefit case due to incomplete information or if a remaining claim that has not yet expired was not taken into account. However, these corrections are usually only administered for the current record, i.e. there is no correction of the complete case. In addition, in individual cases the information on entitlement durations may exceed the individual upper limit according to the statutory requirements. |

### 5.5.14 Start date of unemployment (alo_beg)

| | |
|---|---|
| Variable label | Start date of unemployment |
| Variable name | alo_beg |
| Category | information on employment and benefit receipt |
| Origin | LeH, LHG, ASU, XASU, MTH, XMTH |
| Data type | numerical |
| Detailed description | The variable reports the start date of an uninterrupted sequence of periods of unemployment and is valid at the beginning of the observation.<br><br>The following gaps do not result in an interruption of the period of unemployment:<br>• any gap lasting seven days or less<br>• periods of illness lasting up to 42 days<br><br>The variable is generated within the IEB using multiple data sources. Therefore, it cannot be recreated exactly with the individual data of the LIAB. |

### 5.5.15 Duration of unemployment (alo_dau)

| | |
|---|---|
| Variable label | Duration of unemployment |
| Variable name | alo_dau |
| Category | information on employment and benefit receipt |

| | |
|---|---|
| Origin | LeH, LHG, ASU, XASU, MTH, XMTH |
| Data type | numerical |
| Detailed description | The variable reports the duration (in days) of an uninterrupted sequence of periods of unemployment and is valid at the beginning of the observation.<br><br>The following gaps do not result in an interruption of the period of unemployment:<br>• any gap lasting seven days or less<br>• periods of illness lasting up to 42 days<br><br>However, when calculating the duration these gaps are not considered. Therefore, the duration in this variable may differ from the result of the duration calculation based on the 'Start date of unemployment' (*alo_beg*).<br><br>Prior to 1997, the value 0 does not mean that the individual was not unemployed, as the ASU/XASU sources are not available here.<br><br>The variable is generated within the IEB using multiple data sources. Therefore, it cannot be recreated exactly with the individual data of the LIAB. |

## 5.6 Location data

### 5.6.1 Place of residence - district (Kreis) (wo_kreis)

| | |
|---|---|
| Variable label | Place of residence - district (Kreis) |
| Variable name | wo_kreis |
| Category | location data |
| Origin | BeH, LeH |
| Data type | numerical |
| Hierarchy | federal state \| district |
| Detailed description | For BeH and LeH observations, the place of residence at the district level is only available for the years from 1999 onwards. The variable indicates the district (urban district or rural district) in which the social security contributor lives. The first two digits of the 5-digit district code (Kreisschlüssel) show the code for the federal state (Bundesland, NUTS 1), and positions 1-3 indicate the regional authority (Regierungsbezirk, NUTS 2). Federal states without a regional authority have a '0' in the third position. In the NUTS classification (Nomenclature des unités territoriales statistiques) of the European Union, districts correspond to the level NUTS 3.<br>In the BeH, the place of residence is determined at the end of each year and added consistently to all datasets of a year. In the LeH, the variable contains the place of residence at the beginning of the original time period. This means that the longer the spell lasts, the higher the risk that the place of residence will become obsolete and that the information given for later dates will be incorrect.<br>In order to guarantee consistent regional allocations across the entire observation period, the information on the district was recoded with reference to the territorial allocation of December 31, 2017 for all sources, i.e. in all calendar years, a place of residence is assigned to a district in accordance with the boundaries that the district had on December 31, 2017. As the district boundaries have changed over time, cases would occur in which the district code changes without the individual concerned having relocated if the territorial allocations of the districts were not updated. |

| | |
|---|---|
| Notes on quality | There are inaccuracies in the information provided for some employees with regard to where they live. The reporting requirement does not clarify which residence - main or secondary residence with predominant residence - is to be reported by the employer. In the employment history, this can lead to the impression of "long-distance commuters" between the registered main residence and the place of work, even though the person is in fact employed at the secondary residence, i.e. does not actually commute.<br>In the year 2015, the data show a reduction in the change of residence of approx. 10% to 15% compared to the usual level of the surrounding years, with regional differences occurring. It has not yet been possible to determine the exact reason for this deviation. |
| Anonymisation | Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable 'Place of residence – federal state (Bundesland)' (*wo_bula*) is provided. |

### 5.6.2 Place of residence - federal state (Bundesland) (wo_bula)

| | |
|---|---|
| Variable label | Place of residence - federal state (Bundesland) |
| Variable name | wo_bula |
| Category | location data |
| Origin | BeH, LeH |
| Data type | numerical |
| Hierarchy | federal state \| district |
| Detailed description | This variable is an aggregation of the variable 'Place of residence - district (Kreis)' (*wo_kreis*) to the 16 German federal states.<br>For BeH and LeH observations, the federal state of the place of residence is only available for the years from 1999 onwards. Further information on the district of the place of residence can be found in Section 5.6.1. |

### 5.6.3 Place of residence - employment agency (Arbeitsagentur) (wo_aa)

| | |
|---|---|
| Variable label | Place of residence - employment agency (Arbeitsagentur) |
| Variable name | wo_aa |
| Category | location data |
| Origin | BeH, LeH |
| Data type | numerical |
| Hierarchy | regional directorate \| employment agency |
| Detailed description | From 1999 onwards, this variable contains the agency district of the employment agency that is responsible for the employee's / BA client's place of residence. This information is determined from the residence address. For the LHG and XASU data sources, the place of residence is valid for the period of the original observation. In the case of the LeH, the variable contains the place of residence at the start of the period of unemployment or job search. Accordingly, the longer the spell lasts, the higher the risk that the place of residence will become obsolete and that the information given for later dates will be incorrect. For the BeH and LeH, it is available from 1999 onwards.<br>In order to guarantee consistent regional allocations across the entire observation period, the information on the agency district was recoded to the territorial allocation of December 31, 2017 for all data sources, i.e. in all calendar years, a place of residence is assigned to an agency district in accordance with the boundaries that the agency district had on December 31, 2017. As the boundaries of the agency district have changed over time, cases would occur in which a person's employment |

| | |
|---|---|
| | agency area changes without him/her having relocated if the territorial allocations were not updated.<br>Berlin constitutes a problematic case with regard to updating territorial allocations, however: The boundaries of Berlin's employment agency areas have been changed repeatedly over the years, which could not be fully corrected even by recoding the territorial allocations. Berlin should preferably be analysed as a unit and not separately for East and West, because since the territorial reform of the employment offices in Berlin on July 1, 1997 their clear allocation to East and West is no longer possible. The BA statistics uniformly assigns Berlin to the East. Prior to 1999 the variable contains the employment agency that last processed the case of benefit receipt. It is not determined from the residence address and is only available for the LeH data source. Moreover, it is not possible to update the territorial allocations. This must be taken into account when conducting calculations over time (e.g., regional unemployment figures). |
| Anonymisation | Owing to its particular sensitivity with regard to data privacy, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the area of the regional directorate in which the social security contributor's place of residence is located is shown. |

### 5.6.4 Place of residence - regional directorate (Regionaldirektion) (wo_rd)

| | |
|---|---|
| Variable label | Place of residence - regional directorate (Regionaldirektion) |
| Variable name | wo_rd |
| Category | location data |
| Origin | BeH, LeH |
| Data type | numerical |
| Hierarchy | regional directorate \| employment agency |
| Detailed description | This variable is an aggregation of the variable 'Place of residence: employment agency (Arbeitsangentur)' (*wo_aa*) at the level of the regional directorates. Further information can be found in Section 5.6.3. |

## 5.7 Establishment variables

### 5.7.1 Classification of economic activities 73 (w73_3)

| | |
|---|---|
| Variable label | classification of economic activities 73 |
| Variable name | w73_3 |
| Category | establishment variables |
| Origin | BHP |
| Data type | numerical |
| Hierarchy | division (1-digit code) \| group (2-digit code) \| class (3-digit code) of economic activity |
| Detailed description | This variable indicates the economic activity as a 3-digit code in accordance with the WS73 classification and is available from 1975 up to and including 2002.<br>WS73 stands for the 'Classification of Economic Activities for the Statistics of the Federal Employment Services, edition 1973' ("Klassifikation der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit, Ausgabe 1973"). Using a 3-digit code, the classification distinguishes between 269 classes of economic activity, whereby the first digit of the code defines the division of economic activity of a total of 10, and the first two digits together define the particular group of economic activity of a total of 95. |

| | |
|---|---|
| | Each establishment is only assigned one code. The assignment to the relevant class of economic activity is carried out under consideration of the institutional orientation of the establishment. |

### 5.7.2 Classification of economic activities 93, sub-classes (w93_5)

| | |
|---|---|
| Variable label | classification of economic activities 93, sub-classes |
| Variable name | w93_5 |
| Category | establishment variables |
| Origin | BHP |
| 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 | This variable indicates the economic activity as a 5-digit code in accordance with the WZ93 classification and is available from 1999 up to and including 2003. WZ93 stands for the 'Classification of Economic Activities for the Statistics of the Federal Employment Services, edition 1993' ("Klassifikation der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit, Ausgabe 1993"). The WZ93 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 1 („Nomenclature génerale des activités économiques dans les communautés européennes") which has four levels the first two of which are based on the international standard ISIC Rev. 3 ('International Standard Industrial Classification of All Economic Activities').<br>Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected. |
| Anonymisation | Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, the economic activity is only shown as the 3-digit code (*w93_3*). |

### 5.7.3 Classification of economic activities 93, groups (w93_3)

| | |
|---|---|
| Variable label | classification of economic activities 93, groups |
| Variable name | w93_3 |
| Category | establishment variables |
| Origin | BHP |
| 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 | This variable indicates the economic activity as a 3-digit code in accordance with the WZ93 classification and is available from 1999 up to and including 2003. WZ93 stands for the 'Classification of Economic Activities for the Statistics of the Federal Employment Services, edition 1993' ("Klassifikation der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit, Ausgabe 1993"). The WZ93 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 1 ('Nomenclature génerale des activités économiques dans les communautés européennes') which has four levels the first two of which are based on the international standard ISIC Rev. 3 ('International Standard Industrial Classification of All Economic Activities').<br>Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected. |

### 5.7.4 Classification of economic activities 03, sub-classes (w03_5)

| | |
|---|---|
| Variable label | classification of economic activities 03, sub-classes |
| Variable name | w03_5 |
| Category | establishment variables |
| Origin | BHP |
| 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 | This variable indicates the economic activity as a 5-digit code in accordance with the WZ03 classification and is available from 2003 up to 2008. WZ03 stands for the 'Classification of Economic Activities, Edition 2003' ("Klassifikation der Wirtschaftszweige Ausgabe 2003") of the Federal Statistical Office (eds.). Like the WZ93, the WZ03 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 1 (see description of variables *w93_3*, *w93_5*). The classifications of the economic activity have been updated, but the structure of the WZ93 has been largely retained.<br>Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected. |
| Anonymisation | Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, the economic activity is only shown as the 3-digit code (*w03_3*). |

### 5.7.5 Classification of economic activities 03, groups (w03_3)

| | |
|---|---|
| Variable label | classification of economic activities 03, groups |
| Variable name | w03_3 |
| Category | establishment variables |
| Origin | BHP |
| 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 | This variable indicates the economic activity as a 3-digit code in accordance with the WZ03 classification and is available from 2003 up to 2008. WZ03 stands for the 'Classification of Economic Activities, Edition 2003' ("Klassifikation der Wirtschaftszweige Ausgabe 2003") of the Federal Statistical Office (eds.). Like the WZ93, the WZ03 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 1 (see description of variables *w93_3*, *w93_5*). The classifications of the economic activity have been updated, but the structure of the WZ93 has been largely retained.<br>Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected. |

### 5.7.6 Classification of economic activities 08, sub-classes (w08_5)

| | |
|---|---|
| Variable label | classification of economic activities 08, sub-classes |
| Variable name | w08_5 |
| Category | establishment variables |
| Origin | BHP |
| 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 | This variable indicates the economic activity as a 5-digit code in accordance with the WZ08 classification and is available from 2008 onwards. WZ08 stands for the 'Classification of Economic Activities, Edition 2008' ("Klassifikation der Wirtschaftszweige Ausgabe 2008") of the Federal Statistical Office (eds.). The WZ08 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 2.<br>Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected. |
| Anonymisation | Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise the economic activity is only shown as the 3-digit code (*w08_3*). |

### 5.7.7 Classification of economic activities 08, groups (w08_3)

| | |
|---|---|
| Variable label | classification of economic activities 08, groups |
| Variable name | w08_3 |
| Category | establishment variables |
| Origin | BHP |
| 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 | This variable indicates the economic activity as a 3-digit code in accordance with the WZ08 classification and is available for the years since 2008. WZ08 stands for the 'Classification of Economic Activities, Edition 2008' ("Klassifikation der Wirtschaftszweige Ausgabe 2008") of the Federal Statistical Office (eds.). The WZ08 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 2.<br>Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected. |

### 5.7.8 w73_3 completed by extrapolation/imputation (w73_3_gen)

| | |
|---|---|
| Variable label | w73_3 completed by extrapolation/imputation |
| Variable name | w73_3_gen |
| Category | establishment variables |
| Origin | BHP |
| Data type | numerical |
| Hierarchy | division (1-digit code) \| group (2-digit code) \| class (3-digit code) |
| Detailed description | This variable indicates the economic activity as a 3-digit code in accordance with the WZ73 classification. From 1975 up to and including 2002, the variable contains the original values from *w73_3*. From 2003 onwards, the information is either continued or replaced with the help of recoding tables. Thus the variable provides time-consistent information on the economic activity based on the economic activity classification WS73. A detailed description can be found in Eberle et al. (2011).<br>Further information on the WS73 classification can be found in the description of the variable *w73_3*. |

### 5.7.9 Type of imputation w73_3 (group_w73_3)

| | |
|---|---|
| Variable label | Type of imputation w73_3 |
| Variable name | group_w73_3 |
| Category | establishment variables |
| Origin | BHP |
| Data type | numerical |
| Detailed description | This variable indicates the type of completion for the *w73_3_gen* variable. It reports whether the respective value in *w73_3_gen* is consistent with the original value from *w73_3*, still missing / extrapolated or imputed based on recording tables.<br>A detailed description of the procedure can be found in Eberle et al. (2011). |

### 5.7.10 w93_3 completed by extrapolation/imputation (w93_3_gen)

| | |
|---|---|
| Variable label | w93_3 completed by extrapolation/imputation |
| Variable name | w93_3_gen |
| Category | establishment variables |
| Origin | BHP |
| 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 | This variable indicates the 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).<br>Further information on the WS93 classification can be found in the description of the variable *w93_3*. |

### 5.7.11 Type of imputation w93_3 (group_w93_3)

| | |
|---|---|
| Variable label | Type of imputation w93_3 |
| Variable name | group_w93_3 |
| Category | establishment variables |
| Origin | BHP |
| Data type | numerical |
| Detailed description | This variable indicates the type of completion for the *w93_3_gen* variable. It reports whether the respective value in *w93_3_gen* is consistent with the original value from *w93_3*, still missing / extrapolated or imputed based on recording tables.<br>A detailed description of the procedure can be found in Eberle et al. (2011). |

### 5.7.12 w08_3 completed by extrapolation/imputation (w08_3_gen)

| | |
|---|---|
| Variable label | w08_3 completed by extrapolation/imputation |
| Variable name | w08_3_gen |
| Category | establishment variables |
| Origin | BHP |
| 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 | This variable indicates the economic activity as a 3-digit code in accordance with the WZ08 classification and is available from 2008 onwards. From 2008 up to and including 2017, the variable contains the original values from *w08_3*. Before 2008, the information is either written back 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 WZ08. A detailed description can be found in Eberle et al. (2011).<br>Further information on the WZ08 classification can be found in the description of the variable *w08_3*. |

### 5.7.13 Type of imputation w08_3 (group_w08_3)

| | |
|---|---|
| Variable label | Type of imputation w08_3 |
| Variable name | group_w08_3 |
| Category | establishment variables |
| Origin | BHP |
| Data type | numerical |
| Detailed description | This variable indicates the type of completion for the *w08_3_gen* variable. It reports whether the respective value in *w08_3_gen* is consistent with the original value from *w08_3*, still missing / extrapolated or imputed based on recording tables.<br>A detailed description of the procedure can be found in Eberle et al. (2011). |

### 5.7.14 Year of first appearance (grd_jahr)

| | |
|---|---|
| Variable label | year of first appearance |
| Variable name | grd_jahr |
| Category | establishment variables |
| Origin | BHP |
| Data type | numerical |
| Detailed description | This variable indicates the first appearance of the establishment number in the dataset. If an establishment number in western Germany is only determined for the first time after 1975 (or after 1992 in eastern Germany), this variable could indicate the date when the respective establishment was founded. However, it could also be an establishment that has been in existence for a longer time but has been allocated a new establishment number following a change of owner or a change in the legal form of the establishment (for the allocation of establishment numbers see Bundesagentur für Arbeit 2007, pp. 9-11). It is also possible that the establishment already existed before, but had no employees subject to social security, or from 1999 onwards, no marginal part-time workers.<br>An establishment does not necessarily have to be included in the BHP in the year of its first appearance, as only the key date June 30 is relevant for this inclusion. If the establishment has no employees on June 30 of its year of foundation, it consequently does not appear in the BHP in that year. |

### 5.7.15 First appearance (grd_dat)

| | |
|---|---|
| Variable label | first appearance |
| Variable name | grd_dat |
| Category | establishment variables |
| Origin | BHP |
| Data type | date |

| | |
|---|---|
| Detailed description | This variable indicates the first appearance of the establishment number in the BeH to the day. If an establishment number in western Germany is only determined for the first time after 1975, or after 1992 in eastern Germany, this variable could indicate the date when the respective establishment was founded. However, it could also be an establishment that has been in existence for a longer time but has been allocated a new establishment number following a change of ownership or a change in the legal form of the establishment (for the allocation of establishment numbers see Bundesagentur für Arbeit 2007, pp. 9-11). It is also possible that the establishment already existed before, but had no employees subject to social security, or from 1999 onwards, no marginal part-time workers.<br>An establishment does not necessarily have to be included in the BHP in the year of its first appearance, as only the key date June 30 is relevant for this inclusion. If the establishment has no employees on June 30 of its year of foundation, it consequently does not appear in the BHP in that year. |
| Anonymisation | Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the year when the establishment number first appeared is shown (*grd_jahr*). |

### 5.7.16 Year of last appearance (lzt_jahr)

| | |
|---|---|
| Variable label | year of last appearance |
| Variable name | lzt_jahr |
| Category | establishment variables |
| Origin | BHP |
| Data type | numerical |
| Detailed description | This variable indicates the last appearance of the establishment number in the dataset (see Bender et. al. 1996). If the existence of an establishment number in the BHP already ends before 2017, this could indicate the closure of the establishment. However, other possible reasons for this are an 'arbitrary change of the establishment number following a change of owner or a change in the legal form of the establishment', the 'outsourcing of parts of the firm under a new establishment number' or other administrative changes (see Bender et. al. 1996 or Bundesagentur für Arbeit 2007, pp. 9-11).<br>An establishment does not necessarily have to be included in the BHP in the year of its last appearance, as only the key date June 30 is relevant for this inclusion. If the establishment has no employees on June 30 of its year of closure, it consequently does not appear in the BHP in that year. |

### 5.7.17 Last appearance (lzt_dat)

| | |
|---|---|
| Variable label | last appearance |
| Variable name | lzt_dat |
| Category | establishment variables |
| Origin | BHP |
| Data type | date |

| | |
|---|---|
| Detailed description | This variable indicates the last appearance of the establishment number in the dataset to the day (see Bender et. al. 1996). If the existence of an establishment number in the BHP already ends before 2017, it could indicate the closure of the establishment. However, other possible reasons for this are an 'arbitrary change of the establishment number following a change of ownership or a change in the legal form of the establishment', the 'outsourcing of parts of the firm under a new establishment number' or other administrative changes (see Bender et al. 1996 or Bundesagentur für Arbeit 2007, pp. 9-11).<br>An establishment does not necessarily have to be included in the BHP in the year of its last appearance, as only the key date June 30 is relevant for this inclusion. If the establishment has no employees on June 30 of its year of closure, it consequently does not appear in the BHP in that year. |
| Anonymisation | Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the year when the establishment number last appeared is shown (*lzt_jahr*). |

### 5.7.18 Total number of employees (az_ges)

| | |
|---|---|
| Variable label | no. employees |
| Variable name | az_ges |
| Category | generated establishment variables |
| Origin | BHP |
| Data type | numerical |
| Detailed description | This variable contains the total number of an establishment's employees reported to the social security agencies as of June 30 of a year. Since the introduction of the new notification regulations in 1999, people in marginal part-time employment have also been recorded. Dormant employment relationships (daily wage of zero) are not included. |

### 5.7.19 Number of full-time employees (regular workers + others) (az_vz)

| | |
|---|---|
| Variable label | No. full-time (regular workers + others) |
| Variable name | az_vz |
| Category | generated establishment variables |
| Origin | BHP |
| Data type | numerical |
| Detailed description | This variable contains the number of people in the establishment who are reported on June 30 of a year as full-time employees under the person group codes 101, 140, 143, 105, 106, 112, 113, 114, 118, 119, 120, 149, 201, 203, 205, 999 and YYY. Apprentices, marginally part-time employees and individuals participating in partial retirement schemes are not considered. |

### 5.7.20 Number of employees in marginal part-time employment (az_gf)

| | |
|---|---|
| Variable label | no. marginal part-time workers |
| Variable name | az_gf |
| Category | generated establishment variables |
| Origin | BHP |
| Data type | numerical |
| Detailed description | The number of employees in marginal part-time employment is generated using the person group code – values 109 and 209. This variable has only been contained in |

| | |
|---|---|
| | the dataset since 1999 as it has only been included in the social security notification procedure since that year. |

### 5.7.21 Mean imputed wage all full-time employees (te_imp_mw)

| | |
|---|---|
| Variable label | mean imp. wage all full-time employees |
| Variable name | te_imp_mw |
| Category | generated establishment variables |
| Origin | BHP |
| Data type | numerical |
| Detailed description | This variable contains the mean imputed gross daily wage of the full-time employees in an establishment. It does not include the wages of marginally part-time staff, apprentices or individuals participating in partial retirement schemes.<br>The values are reported in Euros for all years.<br>According to the social security notification regulations, employers must indicate the employee's gross wage subject to social security contributions for a certain period of time (fixed period wage). Until the end of 1998, employers had to report the gross wage subject to social security contributions only. So only wages above the marginal part-time income threshold and below the contribution assessment ceiling were recorded. Since 1999, wages below the marginal part-time income threshold have also been recorded as part of the new notification procedure. Gross wages above the contribution assessment ceiling, however, are still capped.<br>In order to calculate the gross daily wage, the fixed period wage is divided by the number of calendar days in the period. To calculate the mean, these censored wages were imputed (see Section 3.1.3.3 in Schmucker et al. 2018). These data were then aggregated at establishment level. The values are rounded to two decimal places. However, due to the 'storage type' in Stata, additional decimal places are displayed that are not correct. |

### 5.7.22 Place of work - district (Kreis) (ao_kreis)

| | |
|---|---|
| Variable label | Place of work - district (Kreis) |
| Variable name | ao_kreis |
| Category | location data |
| Origin | BHP |
| Data type | numerical |
| Hierarchy | federal state \| district |
| Detailed description | The variable indicates the district (urban district or rural district) 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, NUTS 1), positions 1-3 indicate the regional authority (Regierungsbezirk, NUTS 2). Federal states without a regional authority have a 0 in the third position. In the NUTS classification (Nomenclature des unités territoriales statistiques) of the European Union, districts correspond to the level NUTS 3.<br>In order to guarantee consistent regional allocations across the entire observation period, the information on the district was recoded to the territorial allocation of December 31, 2017, 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 December 31, 2017. 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. |

| | |
|---|---|
| Anonymisation | Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the federal state (*ao_bula*) is shown as regional information. |

### 5.7.23 Place of work - federal state (Bundesland) (ao_bula)

| | |
|---|---|
| Variable label | Place of work - federal state (Bundesland) |
| Variable name | ao_bula |
| Category | location data |
| Origin | BHP |
| Data type | numerical |
| Hierarchy | federal state \| district |
| Detailed description | The variable indicates the federal state in which the establishment is located. This variable is generated from the district code (*ao_kreis*). The first two positions of the district code indicate the federal state (NUTS 1). |

# 6 References


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

Bertat, Thomas / Dundler, Agnes / Grimm, Christopher / Kiewitt, Jochen / Schomaker, Christine / Schridde, Dr. Henning / Zemann, Dr. Christian (2013): Neue Erhebungsinhalte 'Arbeitszeit', 'ausgeübte Tätigkeit' sowie 'Schul- und Berufsabschluss' in der Beschäftigungsstatistik. Methodenbericht, Bundesagentur für Arbeit – Statistik, URL: https://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Methodenberichte/Beschaeftigungsstatistik/Generische-Publikationen/Methodenbericht-Neue-Erbebungsinhalte-Arbeitszeita-ausgeuebte-Taetigkeit-sowie-Schul-und-Berufsabschluss-in-der-Beschaeftigungsstatistik.pdf (accessed: June 12, 2019).

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

Bundesagentur für Arbeit (ed.) (2007): Handbuch für die Betriebsnummernvergabe und –pflege im Rahmen des Meldeverfahren zur Sozialversicherung. Issue December 2007, Nuremberg.

Bundesagentur für Arbeit (2009): Klassifikation der Wirtschaftszweige 1973, Nuremberg. URL: http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Wirtschaftszweige/Klassifikation-der-Wirtschaftszweige-1973-2003/Klassifikationen-der-Wirtschaftszweige-1973-2003-Nav.html (accessed: June 12, 2019)

Bundesagentur für Arbeit (2011): Klassifikation der Berufe 2010. Band 1: Systematischer und alphabetischer Teil mit Erläuterungen, Nuremberg. URL: http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB2010/Printausgabe-KldB-2010-Nav.html (accessed: June 12, 2019).

Bundesanstalt für Arbeit (1988): Klassifikation der Berufe – Systematisches und Alphabetisches Verzeichnis der Berufsbenennung, Nuremberg. URL: http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB1975-1992/KldB1975-1992-Nav.html (accessed: June 12, 2019)

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

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

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.

Eberle, Johanna/ Schmucker, Alexandra (2019): Creating cross-sectional data and biographical variables with the Sample of Integrated Labour Market Biographies 1975-2017 * Programming examples for Stata. FDZ Methodenreport, 04/2019 (en), Nuremberg.

Fischer, Gabriele/ Janik, Florian/ Müller, Dana/ Schmucker, Alexandra (2008): The IAB Establishment Panel – from Sample to Survey to Projection. FDZ Methodenreport, 01/2008 (en), Nuremberg.

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, vol. 126, issue 3, pp. 405-436.

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

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

Paulus, Wiebke/ Matthes, Britta (2013): The German Classification of Occupations 2010 – Structure, Coding and Conversion Table. FDZ-Methodenreport, 08/2013 (en), Nuremberg.

Schmucker, Alexandra/Eberle, Johanna/ Ganzer, Andreas/Stegmaier, Jens/ Umkehrer, Matthias (2018): Establishment History Panel 1975-2016. FDZ-Datenreport, 01/2018 (en), Nuremberg.

Schmucker, Alexandra/ Seth, Stefan/ Ludsteck, Johannes/ Eberle, Johanna/ Ganzer, Andreas (2016): Establishment History Panel 1975-2014. FDZ Datenreport, 03/2016 (en), Nuremberg.

Statistisches Bundesamt (2002): Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ 93), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-1993.html (accessed: June 12, 2019).

Statistisches Bundesamt (2003): Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftklassifikationen/Content75/KlassifikationWZ2003.html (accessed: June 12, 2019).

Statistisches Bundesamt (2008): Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html (accessed: June 12, 2019).

Statistisches Bundesamt (2019): Staats- und Gebietssystematik, Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluessel_pdf.pdf?__blob=publicationFile&v=5 (accessed: June 12, 2019).

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

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.

# 7 Appendix

## 7.1 Frequency tables

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

## 7.2 List of abbreviations

| | | |
|---|---|---|
| **AA** | Agentur für Arbeit / Arbeitsamt | employment agency / employment office |
| **ALG** | Arbeitslosengeld | unemployment benefit |
| **ALH** | Arbeitslosenhilfe | unemployment support |
| **ARGE** | Arbeitsgemeinschaft | cooperation of employment agencies and municipalities |
| **ASU** | Arbeitsuchende-Historik | Jobseeker History |
| **A2LL** | Arbeitslosengeld II – Leistungen zum Lebensunterhalt | unemployment benefit II - benefits to secure a livelihood |
| **BA** | Bundesagentur für Arbeit | Federal Employment Agency |
| **BeH** | Beschäftigten-Historik | Employee History |
| **BfA** | Bundesversicherungsanstalt für Angestellte | Federal Social Insurance Office for Salaried Employees |
| **BG** | Bedarfsgemeinschaft | Benefit unit |
| **BHP** | Betriebs-Historik-Panel | Establishment History Panel |
| **BMAS** | Bundesministerium für Arbeit und Soziales | Federal Ministry of Labour and Social Affairs |
| **coArb** | Computerunterstützte Arbeitsvermittlung (operatives Verfahren zur Verwaltung der Vermittlung (Altverfahren)) | computer-aided job placement (procedure for the administration of job placements – old procedure) |
| **DEÜV** | 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 |
| **DEVO** | 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 |
| **DIM** | Daten- und IT-Management | Data and IT Management |
| **DÜVO** | 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 |
| **EDV** | Elektronische Datenverarbeitung | Electronic data processing |
| **FDZ** | Forschungsdatenzentrum | Research Data Centre |
| **FELEG** | Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit | Act on the Support in Case of Termination of Farming Activities |

| | | |
|---|---|---|
| **gAw** | Träger mit getrennter Aufgabenwahrnehmung | Municipalities exercising their duties separately |
| **gE** | Gemeinsame Einrichtung | Joint facility |
| **gT** | Getrennte Trägerschaft | Separated responsibilities |
| **IAB** | Institut für Arbeitsmarkt- und Berufsforschung | Institute for Employment Research |
| **IEB** | Integrierte Erwerbsbiographien | Integrated Employment Biographies |
| **ISIC** | International Standard Industrial Classification of All Economic Activities | International Standard Industrial Classification of All Economic Activities |
| **KldB** | Klassifikation der Berufe | Classification of occupations |
| **LeH** | Leistungsempfängerhistorik | Benefit Recipient History |
| **LHG** | Leistungshistorik Grundsicherung | Unemployment Benefit II Recipient History |
| **LIAB** | Linked-Employer-Employee-Daten des IAB | Linked Employer-Employee Data of the IAB |
| **LM** | Längsschnittmodell | Longitudinal Model |
| **LVA** | Landesversicherungsanstalt | Land Social Insurance Office |
| **MTH** | Maßnahmeteilnehmer-Historik | Participants-in-Measures History File |
| **NACE** | Nomenclature génerale des activités économiques dans les communautés européennes | Nomenclature génerale des activités économiques dans les communautés européennes |
| **NUTS** | Nomenclature des unités territoriales statistiques | Nomenclature des unités territoriales statistiques |
| **QM** | Querschnittmodell | Cross-sectional model |
| **SGB** | Sozialgesetzbuch | German Social Code |
| **UHG** | Unterhaltsgeld | subsistence payment |
| **VerBIS** | Vermittlungs- und Beratungsinformationssystems | Information System for Placement and Counselling |
| **XASU** | Arbeitsuchenden-Historik aus XSozial-BA-SGB II | Jobseeker History from XSozial-BA-SGB II |
| **XLHG** | Leistungshistorik Grundsicherung aus XSozial-BA-SGB II | Unemployment Benefit II Recipient History from XSozial-BA-SGB II |
| **zkT** | Zugelassene kommunale Träger | Authorised municipalities |

# Imprint

**FDZ-Datenreport 05|2019 EN**

**Date of publication**

October 01, 2019

**Publisher**

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



**Download**

http://doku.iab.de/fdz/reporte/2019/DR_05-19_EN.pdf

**Documentation version**

LIAB_LM_7517_EN_v1_dok1, DOI: 10.5164/IAB.FDZD.1905.en.v1

**Dataset version**

LIAB LM 7517 v1, DOI: 10.5164/IAB.LIABLM7517.de.en.v1

**All publications in the series 'FDZ-Datenreport' can be downloaded from**

https://fdz.iab.de/en/FDZ_Publications/FDZ_Publication_Series/FDZ-Datenreporte.aspx

**Website**

http://fdz.iab.de

---

**Corresponding author**

Matthias Umkehrer
Phone: 0911 179-6211
Email: Matthias.Umkehrer@iab.de