



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

07|2019 EN Linked Personnel Panel survey data linked with
administrative data of the IAB (LPP-ADIAB) 1975–2017

Jan Mackeben, Kevin Ruf, Tobias Haepf, Stefanie Wolter, Philipp Grunau



Bundesagentur für Arbeit

Linked Personnel Panel survey data linked with administrative data of the IAB (LPP-ADIAB) 1975–2017

Autoren:

Jan Mackeben (IAB), Kevin Ruf (IAB),

Tobias Haepf (IAB), Stefanie Wolter (IAB), Philipp Grunau (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 can be used to prepare evaluations.

Contents

1	Introduction and Short Description	7
1.1	Introduction	7
1.2	Data use.....	8
1.2.1	Data access.....	8
1.2.2	Data management	8
1.2.3	Data combination	10
1.3	Changes as compared to previous versions	12
1.4	Outline	14
1.5	List of variables	17
1.6	Volume structure.....	21
2	Data sources	21
2.1	LPP Establishment Survey.....	21
2.2	LPP Employee Survey	22
2.3	Survey Data from the IAB Establishment Panel.....	22
2.4	Administrative Data from the IEB.....	23
2.4.1	Employee History (BeH)	23
2.4.2	Benefit Recipient History (LeH).....	24
2.4.3	Unemployment Benefit II Recipient History (LHG)	24
2.4.4	Jobseeker Histories (ASU/XASU).....	25
3	Data preparation	25
3.1	LPP survey data.....	25
3.1.1	Dataset descriptions	25
3.1.2	Data preparation changes	25
3.2	Administrative data from the IEB.....	26
3.2.1	Corrections and validation procedures.....	26
3.2.2	Employee History (BeH)	26
3.2.3	Benefit Recipient History (LeH).....	27
3.2.4	Unemployment Benefit II Recipient History (LHG)	27
3.2.5	Jobseeker Histories (ASU/XASU).....	27
3.2.6	SGB II anonymization	28
3.2.7	Episode splitting.....	28
3.2.8	Sampling procedure.....	29
3.2.9	Missing values.....	30
4	Data quality and problems	30
4.1	Integrated Employment Biographies (IEB) in General.....	30
4.1.1	Gaps in employment histories	30

4.1.2	Introduction of SGB II and subsequent institutional changes.....	31
4.2	Employee History (BeH).....	32
4.3	Benefit Recipient History (LeH)	34
4.4	Unemployment Benefit II Recipient History (LHG)	35
4.5	Jobseeker Histories (ASU/XASU)	36
4.5.1	ASU.....	36
4.5.2	XASU.....	37
5	Description of variables	38
5.1	Identifiers	38
5.1.1	Individual ID (persnr).....	38
5.1.2	IAB-Establishment-Panel-ID (idnum)	38
5.1.3	Establishment ID (betnr)	38
5.2	Variables from survey para data	39
5.2.1	Interview date (intdat)	39
5.2.2	Weighting factor LPP Employee Survey (wave 1).....	40
5.2.3	Weighting factor LPP Employee Survey: Cross (wave 2).....	40
5.2.4	Weighting factor LPP Employee Survey: Cross (mover, wave 2)	40
5.2.5	Weighting factor LPP Employee Survey: Cross (mover, exits, wave 2).....	40
5.2.6	Weighting factor LPP Employee Survey: Panel (wave 2)	40
5.2.7	Weighting factor LPP Employee Survey: Panel (mover, wave 2).....	41
5.2.8	Weighting factor LPP Employee Survey: Panel (mover, exits wave 2)	41
5.2.9	Weighting factor LPP Employee Survey: Cross (wave 3).....	41
5.2.10	Weighting factor LPP Employee Survey: Cross (mover, wave 3)	41
5.2.11	Weighting factor LPP Employee Survey: Cross (mover, exits, wave 3).....	42
5.2.12	Weighting factor LPP Employee Survey: Panel (wave 3)	42
5.2.13	Weighting factor LPP Employee Survey: Panel (mover, wave 3).....	42
5.2.14	Weighting factor LPP Employee Survey: Panel (mover, exits wave 3)	42
5.3	Generated technical variables.....	42
5.3.1	Observation counter per person (spell)	42
5.3.2	Source of spell (quelle).....	43
5.3.3	Year (jahr)	43
5.4	Period of validity	43
5.4.1	Original start date (begorig)	43
5.4.2	Original end date of observation (endorig).....	44
5.4.3	Start date of split episode (begepi)	44
5.4.4	End date of split episode (endeipi)	44
5.5	Personal information	45
5.5.1	Gender (frau)	45
5.5.2	Year of birth (gebjahr)	45
5.5.3	Month of birth (gebmon).....	45
5.5.4	Nationality (nation).....	45

5.5.5	Nationality, grouped (nation_gr).....	46
5.5.6	Marital status (famst)	46
5.5.7	Number of children (kind).....	46
5.5.8	Vocational training (ausbildung)	47
5.5.9	Vocational training (imputed) (ausbildung_imp)	50
5.5.10	School leaving qualification (schule)	50
5.6	Information on employment, benefit receipt and job search.....	51
5.6.1	Daily wage, daily benefit rate (tengtelt)	51
5.6.2	Occupation - current/most recent (KldB 1988) (beruf)	52
5.6.3	Occupational group - current/most recent (KldB 2010) (beruf2010_3)	53
5.6.4	Occupational sub-group - current/most recent (KldB 2010) (beruf2010_4)	54
5.6.5	Level of requirement - current/most recent job (KldB 2010) (niveau)	55
5.6.6	Part-time (teilzeit)	55
5.6.7	Employment status (erwstat)	56
5.6.8	Transition zone (gleitz)	58
5.6.9	Temporary agency work (leih)	58
5.6.10	Fixed-term contract (befrist).....	58
5.6.11	Reason of cancellation/notification/termination (grund).....	58
5.6.12	Employment status prior to job search (estatvor)	60
5.6.13	Employment status after job search (estatnach).....	60
5.6.14	Client profile (profil)	60
5.6.15	Type of termination of last job (art_kuend)	61
5.6.16	Desired working hours of the job sought (arbzeit).....	61
5.6.17	Residual claim/planned duration (restanspruch).....	61
5.6.18	Type of provider (traeger)	62
5.6.19	Start date of unemployment (alo_beg).....	63
5.6.20	Duration of unemployment (alo_dau)	63
5.7	Location data	63
5.7.1	Place of residence - district (Kreis) (wo_kreis)	63
5.7.2	Place of residence - federal state (Bundesland) (wo_bula).....	64
5.7.3	Place of residence - employment agency (Arbeitsagentur) (wo_aa)	65
5.7.4	Place of residence - regional directorate (Regionaldirektion) (wo_rd)	65
5.8	Establishment variables	66
5.8.1	Classification of economic activities 73 (w73_3)	66
5.8.2	Classification of economic activities 93, sub-classes (w93_5)	66
5.8.3	Classification of economic activities 93, groups (w93_3)	66
5.8.4	Classification of economic activities 03, sub-classes (w03_5)	67
5.8.5	Classification of economic activities 03, groups (w03_3)	67
5.8.6	Classification of economic activities 08, sub-classes (w08_5)	68
5.8.7	Classification of economic activities 08, groups (w08_3)	68
5.8.8	w73_3 completed by extrapolation/imputation (w73_3_gen)	68
5.8.9	Type of imputation w73_3 (group_w73_3)	69
5.8.10	w93_3 completed by extrapolation/imputation (w93_3_gen)	69

5.8.11 Type of imputation w93_3 (group_w93_3)	69
5.8.12 w08_3 completed by extrapolation/imputation (w08_3_gen)	70
5.8.13 Type of imputation w08_3 (group_w08_3)	70
5.8.14 Year of first appearance (grd_jahr).....	70
5.8.15 First appearance (grd_dat)	71
5.8.16 Year of last appearance (lzt_jahr).....	71
5.8.17 Last appearance (lzt_dat)	71
5.8.18 Total number of employees (az_ges)	72
5.8.19 Number of full-time employees (regular workers + others) (az_vz).....	72
5.8.20 Number of employees in marginal part-time employment (az_gf)	72
5.8.21 Mean imputed wage all full-time employees (te_imp_mw).....	73
5.8.22 Place of work - district (Kreis) (ao_kreis).....	73
5.8.23 Place of work - federal state (Bundesland) (ao_bula)	74
6 References.....	75
7 Appendix	78
7.1 Frequency tables.....	78
7.2 List of abbreviations	78
8 Imprint	0

Zusammenfassung

Dieser Datenreport beschreibt die verknüpften Daten der Befragung „Linked Personnel Panel“ (LPP) mit administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) von 1975 bis 2017.

Abstract

This data report describes the Linked Personnel Panel survey data linked with administrative data of the IAB (LPP-ADIAB) 1975–2017.

Keywords

German administrative micro data, labour market data, data manual

Acknowledgements

The project is financed by the Federal Ministry of Labour and Social Affairs (BMAS) and the Institute for Employment Research (IAB). We would like to thank our cooperation partners, the University of Cologne, the Eberhardt Karls, the University Tübingen and the Centre for European Economic Research (ZEW) as well as the survey institutes infas and Kantar Public. We would like to thank our colleagues in the Research Data Centre (FDZ) 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 04/2018 (authors: Mackeben, Ruf, Grunau, Wolter) with additional inputs from FDZ-Datenreport 02/2019 (authors: Antoni, Schmucker, Seth, vom Berge)

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 Short Description

1.1 Introduction

This data report details the linked survey data of the first three survey waves of the Linked Personnel Panel (LPP) with selected administrative data by the Federal Employment Agency.

This FDZ data report alone is not enough, however, to work with LPP-ADIAB 7517 data; it is instead meant as a supplement to the existing documentation concerning the LPP, the IAB Establishment Panel, and the Establishment History Panel (BHP).

For that reason, please also use the data and method reports on the LPP and the IAB Establishment Panel available on the FDZ website¹.

For information on the LPP, the survey method, and variable descriptions please refer to Mackeben et al. (2018). The first three LPP establishment and employee survey waves from the “Quality of work and economic success” project are the starting point of the LPP-ADIAB. The “Quality of work and economic success: Longitudinal analysis in German businesses” project is a research cooperation between the Institute for Employment Research (IAB), the University of Tübingen, the University of Cologne, and the Centre for European Economic Research (ZEW) funded by the Federal Ministry of Labour and Social Affairs (BMAS) and the IAB. The first step in all waves of the study was to interview establishments with at least 50 employees subject to social security contributions which already formed part of the IAB Establishment Panel, followed by a survey among randomly selected employees of these establishments. For more information on the definition of the population, the sampling process, and the survey procedure, please refer to the published LPP data and method reports².

The information on all establishments ever interviewed for the LPP, which has been gathered during the IAB Establishment Panel, forms the second part of this linked data product.

The administrative individual-level data—forming the third part of this data product—stem from the Integrated Employment Biographies (IEB) of the IAB. The IEB comprises all individuals who showed 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) or at an institution responsible for implementing SGB II as a jobseeker (recorded from 1997 onwards)
- participation in an employment or training measure (recorded from 2000 onwards)

¹ FDZ-Homepage: https://fdz.iab.de/de/FDZ_Overview_of_Data.aspx

² Daten- und Methodenreporte zum LPP: https://fdz.iab.de/de/Integrated_Establishment_and_Individual_Data/lpp/LPP1617.aspx

These data, which come from different sources, are merged in the IEB and the statuses are depicted exact to the day. While the information on employment covered by social security and marginal part-time employment stems from the Employee History (BeH), the reception 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 (Leistungshistorik Grundsicherung - LHG). The Jobseeker Histories (Arbeit-suchendenhistoriken - ASU and XASU) are the data source for the periods of job search recorded by the BA or by municipal institutions responsible for implementing SGB II, while participation in employment and training measures is recorded in the Participants-in-Measures History Files (Maßnahmeteilnahmehistoriken – MTH). Due to the content orientation of the linked data set, the information from the Participants-in-Measures History (MTH) is not contained in the LPP-ADIAB 7517.

The LPP-ADIAB standard version includes the three data sources listed above and may be extended by adding administrative information on the surveyed establishments from the Establishment History Panel (BHP) of the IAB (Schmucker et al., 2018) upon reasonable request.

This data report is structured as follows: chapter 1 contains the introduction and also, among other things, information on data access as well as a short description of the data, the quantity structure, and a list of attributes. A description of the individual data sources can be found in Chapter 2. Data preparation and quality as well as linking are detailed in Chapters 3 and 4, and the individual attributes are described in Chapter 5.

1.2 Data use

1.2.1 Data access

The LPP-ADIAB 7517 data is only available for on-site use at an FDZ location and via remote data processing. We recommend an on-site visit before. In order to be able to use the data, it is first 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 website³.

1.2.2 Data management

In order to save storage space for data maintenance, the LPP-ADIAB 7517 is stored in several files. These files are classified by survey waves as well as by establishment and employee survey. The various files may be merged with the help of the identifiers contained within. Table 1 gives an overview of the files and their respective identifiers.

The LPP survey data serve as the starting point. They contain the attributes of the establishments and their employees gathered during the interviews. While the employee interviews contain the artificial individual ID and the artificial establishment ID as identifiers, the artificial establishment ID and ID number are contained in the establishment survey.

³ Data access and data use: https://fdz.iab.de/de/FDZ_Data_Access/FDZ_On-Site_Use.aspx

The complete waves from 1993 to 2017 were provided for the IAB Establishment Panel. The administrative data of the LPP-ADIAB 7517 consist of the administrative person file. It contains identifiers (system-free person ID and system-free establishment number), personal characteristics, information on employment, benefits received and job search, characteristics on place of residence and technical characteristics. On request, the administrative data of the BHP can also be added to the data of the LPP employers. The link is made here via the system-free employer number and the year. Table 1: LPP-ADIAB 7517 data and identifiers contained therein provides an overview of the files and link variables.

Table 1: LPP-ADIAB 7517 data and identifiers contained therein

Data set	Identifiers
LPP	
lpp-adiab_7517_employer_1213_v1.dta	idnum, betnr
lpp-adiab_7517_employer_1415_v1.dta	idnum, betnr
lpp-adiab_7517_employer_1617_v1.dta	idnum, betnr
lpp-adiab_7517_employee_1213_v1.dta	idnum, persnr, betnr
lpp-adiab_7517_employee_1415_v1.dta	idnum, persnr, betnr
lpp-adiab_7517_employee_1617_v1.dta	idnum, persnr, betnr
IAB-Betriebspanel	
iabbp_1993.dta	idnum
...	...
iabbp_2017.dta	idnum
IEB	
lpp-adiab_7517_v1.dta	betnr, persnr
BHP	
lpp-adiab_7517_bhp_basis_v1.dta	betnr
lpp-adiab_7517_bhp_1975_v1.dta	betnr
...	...
lpp-adiab_7517_bhp_2017_v1.dta	betnr
lpp-adiab_7517_bhp_inflow_v1.dta	betnr
lpp-adiab_7517_bhp_outflow_v1.dta	betnr
lpp-adiab_7517_bhp_entry_v1.dta	betnr
lpp-adiab_7517_bhp_exit_v1.dta	betnr

Certain variables, which make it possible to identify individuals or establishments, 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. These 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 (grd_dat)
- date of last appearance (lzt_dat)

As a rule, these sensitive characteristics are already contained in the data in a coarse version. The listed sensitive characteristics are therefore only made available if the information in the coarsened characteristic is not sufficient to achieve the research objective. In addition, the extension modules “Employee Fluctuation” (inflow/outflow) and “Establishment Dynamics” (start-ups/close-downs) may be requested separately giving reasons. Please refer to https://fdz.iab.de/de/FDZ_Establishment_Data/Establishment_History_Panel.aspx for more information on the BHP.

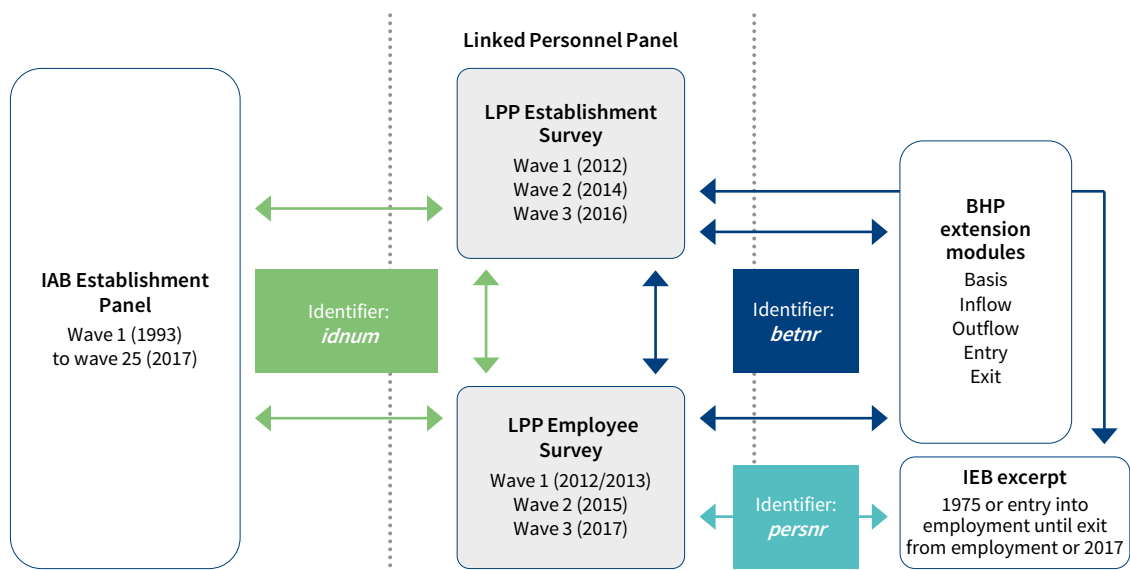
1.2.3 Data combination

The starting point of the linking between survey and process data are the establishments that participated in the two-tier LLP survey and their employees. The establishments were interviewed in a separate additional survey after being interviewed for the IAB Establishment Panel. Afterwards, a sample of their employees was interviewed as well. The employees were asked to give their consent to associate their interview data with the process-produced data about them stored at the BA for research purposes. The question was phrased as follows:

“In order to avoid having to ask you to state your entire occupational history again in the next interview, and not to draw the next interview out any more than necessary, we would like to include excerpts from other data in the evaluation of this survey. This data is available at the Institute for Employment Research in Nuremberg. It includes, for example, information on previous employment times. However, your consent to include this data is required. I kindly ask you to grant your permission to merge these data with the interview data pursuant to the data privacy act. When this information is evaluated, strict compliance with all data protection regulations is guaranteed by all means. Your consent is absolutely voluntary. You may also revoke it at any time.”

While the LPP data include all surveyed employees the LPP-ADIAB includes only those surveyed individuals with linkage consent. Figure 1: Data organisation of the LPP-ADIAB 7517 provides an overview of the entire data organisation of the LPP-ADIAB 7517. The individual components may be merged using the respective common identifiers.

Figure 1: Data organisation of the LPP-ADIAB 7517



The following code can be used to link the establishment and employee data of the second wave in Stata 14.

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, data may be linked using the “merge” command and the year as well as the artificial establishment ID in connection with the respective path (see Box 4).

Box 1 Example code to link establishment and employee data of the second wave in Stata 14

```

use lpp-adiab_7517_employer_1415_v1.dta, clear
merge 1:m idnum using lpp-adiab_7517_employee_1415_v1.dta
  
```

In addition, the IAB Establishment Panel, which is the basis of the LPP establishment sample, is provided for linking in the LPP-ADIAB 7517. The LPP establishment survey and the IAB Establishment Panel are also linked using the establishment ID *idnum*.

Box 2 Example code to link the second wave LPP Establishment Survey with the IAB Establishment Panel in Stata 14

```
use lpp-adiab_7517_employer_1415_v1.dta, clear
merge 1:1 idnum using iabbp_2014_v1.dta
```

The LPP-ADIAB offers the option to merge employment biographies from administrative individual-level data with the employee survey data from the LPP. The individual ID persnr is used to facilitate the linkage.

Box 3 Example code to link the second-wave LPP Employee Survey with administrative individual-level data in Stata 14

```
use lpp-adiab_7517_employer_1415_v1.dta, clear
merge 1:m persnr using lpp-adiab_7517_v1.dta
```

Box 4 Example code to link individual-level data with the basic establishment file in Stata 14

```
use lpp-adiab_7517_v1.dta
gen int jahr = year(begepi)
sort betnr jahr
merge m:1 betnr jahr using lpp-adiab_7517_bhp_basis_v1.dta
```

In order to help users to create panel data sets, Stata programmes are provided in the Tools section on the FDZ website⁴. These programmes are available for the LPP establishment and employee surveys and contain the respective variables contained in both waves. The programme also includes a code for the transfer of attributes for the employee data set. This code transfers answers to questions that only first-time respondents were asked, such as questions about personality or migration background, from the first to the following waves.

1.3 Changes as compared to previous versions

Compared to the previous version of the LPP-ADIAB (7514), the content has been changed as specified below:

- Extension of the accurate Integrated Employment Biographies (IEB) to include all employees of the establishments in the LPP Establishment Survey

⁴ Stata programmes are located at https://fdz.iab.de/de/Integrated_Establishment_and_Individual_Data/lpp/lpp-adiab.aspx

- Extension of the administrative establishment data (BHP) to include all establishments where the employees included in the LPP Employee Survey ever worked
- Extension of the mergeable establishment panel to include all available waves and establishments (1993–2017)

In addition to these changes in content, the adjustment of the weighting factors of the employee surveys described below was necessary. In their first interview, all persons were asked whether they agreed to a link to your survey data. Panel cases who rejected this, were asked again in the follow-up interview. All persons who refused at the end of their last participation are not part of the LPP-ADIAB 7517. Since the original weights were formed on the basis of all persons interviewed in the LPP, an adjustment of the weighting factors was necessary.

As a first step the connection between all variable and the merging question were examined with graphs, cross tables and statistical tests suitable for the various scale levels. In the next step, the data set was prepared for a multiple logistic regression. During this step multiple imputations were calculated for some independent variables. Ultimately, for each case, the linkage probabilities were calculated using a logistic regression model and multiplied by the original weights as reciprocal values. Finally, all weights were normalized to case numbers. The adjusted weights are saved with the suffix “_ADIAB” (see Section 5.2)

Moreover, the following changes in the IEB were adopted to the LPP-ADIAB 7517:

- Correction of incorrect 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.
- Variables included: The LPP-ADIAB 7517 offers a number of additional variables compared to the LPP-ADIAB 7514. The variable "Vocational training (imputed)" (ausbildung_imp, see Section 5.5.9) has been added. This variable contains imputed data for observations with originally missing information on the level of vocational training. Due to increasing inquiries, LPP-ADIAB 7517 now also contains the sensitive variable "Month of birth" (gebmon, see Section 5.5.3). In addition, a time-consistent version of the WZ08 industry classification was generated by extrapolation and imputation. Further information can be found in sections 5.8.11 und 5.8.12. Moreover, changes in the variables included in the LPP employee survey were also implemented. In addition to adding the indicators idnum, betnr and persnr, the interview date has also been included in the LPP employee survey data. This variable can be used to identify the exact employment spell the interview is referring to.

Table 2 provides an overview of the changes in the canon of variables across all sources.

Table 2: New variables in the LPP-ADIAB 7517

Variable	Explanation
Vocational training (imputed) (ausbildung_imp)	Newly included, but with fewer categories than the variable “Vocational training” (ausbildung)
Month of birth (gebmon)	Newly included as a sensitive variable
w08_3 completed by extrapolation/imputation (w08_3_gen)	Newly added to the basic establishment file
Type of completion w08_3 (group_w08_3)	Newly included in Basis Establishment File
Interview date (intdat)	Newly included in LPP Employee Survey

1.4 Outline

Table 3: Outline

Content characteristics	
Topics/ groups of variables	<p>Establishment characteristics from survey (LPP): HR planning and recruitment, HR development, remuneration structure, commitment, values and corporate culture, structural features</p> <p>Employee variables from survey (LPP): Personal characteristics, employment, HR development working conditions and workload, remuneration, commitment, values and corporate culture, personality and attitudes, health, sociodemographic characteristics</p> <p>Establishment characteristics from survey (IAB Establishment Survey) Core modules Employment development, business policy and business development, investments, innovations, public funding, workforce structure, vocational training and apprenticeship training, fluctuations, personnel search, wages and salaries, operational working hours, further training and education, general information about the establishment Exchangeable modules/One-time focus topics (selection): Main focus 2010: Economic and financial crisis; agency workers, short time working and engagement abroad 2015: Minimum wage, older employees, further training</p> <p>Employee History (Beschäftigtenhistorik - BeH): Annual notifications and end-of-employment notifications submitted to the social security agencies for employees covered by social security and employees in marginal part-time employment.</p> <p>Benefit Recipient History (Leistungsempfängerhistorik - LeH):</p>

	<p>Information on benefit receipt in accordance with Social Code Book III (SGB III) for recipients of unemployment benefit, unemployment assistance and maintenance allowance.</p> <p>Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung - LHG): Data on individuals in receipt of basic social security benefits in accordance with Social Code Book II (SGB II) (Types of institution: cooperation of employment agencies and municipalities/joint facilities, separated responsibilities/municipalities exercising their duties separately, authorised municipalities)</p> <p>Jobseeker History (Arbeitsuchendenhistorik - ASU): Information on job search activities that are recorded in BA procedures</p> <p>Jobseeker History from XSozial-BA-SGB II (Arbeitsuchendenhistorik aus XSozial-BA-SGB II - XASU): Information on job search activity reported to the BA by authorised municipalities via the transmission standard XSozial-BA-SGB II to the BA by authorised municipalities.</p>
Data units	<p>Establishment survey: Establishments with more than 50 employees subject to social insurance contributions (Reference date: 30 June of the previous year)</p> <p>Employee survey: Employees subject to social insurance contributions (Reference date panel cases: 31/12/2011 (w1), 31/12/2013 (w2); Refresher: 31/12/2013 (w2); 31/12/2015 (w3))</p>
Number of cases	<p>Survey data: Wave 1: 6.576 Individuals, 1.219 Establishments Wave 2: 6.460 Individuals, 771 Establishments Wave 3: 6.180 Individuals, 846 Establishments</p> <p>Administrative data: 9.718 Individuals 371,554 Original observations of LPP employees 461,188 Non-overlapping observations (after episode splitting) Thereof: 12.035 LPP- Individuals 510.074 Original observations of LPP employees 635.515 Non-overlapping observations) of LPP- employees (after episode splitting) 41.988 Establishments (BHP Basis)</p>
Period covered	<p>Establishment survey LPP: Wave 1: June 2012 to October 2012 Wave 2: June 2014 to October 2014 Wave 3: June 2016 to November 2016</p> <p>Employee survey LPP: Wave 1: November 2012 to February 2013 Wave 2: March 2015 to August 2015 Wave 3: February 2017 to July 2017</p> <p>Administrative data (time period for each data source): BeH: 01.01.1975 - 31.12.2017 (2014: 36-months-file</p>

	<p>2015 and 2016: 18-months-file, 2017: 6- months-file)</p> <p>LeH: 01.01.1975 - 31.12.2017 ASU: 01.01.1997 - 31.12.2017 LHG: 01.01.2005 - 31.12.2017 XASU: 01.01.2005 - 31.12.2017</p>
Time reference	<p>Surveys: Time of interview</p> <p>Administrative Individual data: Exact to the day</p> <p>Administrative Employer data: Reference date 30 June</p>
Regional structure	<p>Establishment survey:</p> <ul style="list-style-type: none"> Northern region: Lower Saxony, Schleswig-Holstein, Hamburg, Bremen Eastern region: Berlin, Brandenburg, Mecklenburg-Western Pomerania, Saxony, Saxony-Anhalt, Thuringia Southern region: Bavaria, Baden-Wuerttemberg Western region: North Rhine-Westphalia, Hesse, Saarland, Rhineland-Palatinate <p>Employee survey: German federal states (Bundesländer)</p>
Methodological characteristics	
Survey Design	<p>LPP Survey data: <i>Establishment survey:</i> Sample of establishments with more than 50 employees subject to social insurance who also participated in the IAB establishment panel. The stratification was based on company size, sector and region. <i>Employee survey:</i> Sample of employees from establishments included in the LPP establishment survey, stratified by establishment size.</p> <p>IAB Establishment Panel: Representative sample of establishments, stratified by establishment size, sector and federal state</p> <p>Administrative data: Identification of all employees working in LPP establishments in the Integrated Employment Biographies of the IAB. Identification of the establishments surveyed in the IAB's Establishment History Panel.</p>
Institutions involved	<p>LPP Survey data: Institute for Employment Research (IAB), Centre for European Economic Research (ZEW), University Cologne, Federal Ministry of Labour and Social Affairs (BMAS), TNS Infratest Sozialforschung, infas Institute for Applied Social Sciences</p> <p>IAB Establishment Panel: Contracting authorities: Research department "Establishments and Employment" of the IAB, administrations of the federal states or regional directorates Implementation: Kantar Public in cooperation with the Institute for Socioeconomic Structure Analysis GmbH (SÖSTRA)</p> <p>Administrative data:</p>

	Social security agencies, Federal Employment Agency (Bundesagentur für Arbeit), municipal institutions
Frequency of data collection	Survey data: Biennial survey Administrative data: Ongoing
File format	Stata
File architecture	The data is stored in separate files. The employee and employer information from the LPP survey is contained in three files each. The waves of the employee survey can be extended by the administrative individual data of the IEB. The IAB-establishment panel is divided into 25 annual files. Given a justified application, additional files with establishment information regarding the surveyed LPP establishments can be provided from the Employer History Panel (BHP).
Data access	On-site usage at the FDZ of the BA at the IAB and subsequent remote data access
Degree of anonymisation	Weakly anonymous
Sensitive variables	Survey data: None Administrative individual data: Month of birth (gebmon), Nationality (nation), occupational sub-group (beruf2010_4), place of residence: employment agency (wo_aa), place of residence: district (Kreis) (wo_kreis), month of birth (gebmon) Administrative establishment data: place of work: district (Kreis) (ao_kreis), economic activity - sub-class of economic activity (five-digit code) (w93_5), economic activity - sub-class of economic activity (five-digit code) (w03_5), economic activity - sub-class of economic activity (five-digit code) (w08_5), first appearance of establishment (grd_dat), last appearance of establishment (lzt_dat)
Citation of data and data documentation	Data: LPP survey data linked with administrative data from the IAB (LPP-ADIAB) 1975–2017 v1 Data documentation: Jan Mackeben, Kevin Ruf, Tobias Haepf, Stefanie Wolter, Philipp Grunau (2019): LPP—Linked Personell Panel Survey data linked with administrative data of the IAB (LPP-ADIAB 7517) 1975–2014. FDZ-Data Report, 07/2019 (en), Nuremberg. DOI: 10.5164/IAB.FDZD.1907.en.v1

1.5 List of variables

The overview of variables in Table 5: List of variables with degree of completeness 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. Due to the high number of variables in the survey data, the variables from the LPP survey and the IAB Establishment Panel will only be listed in the corresponding original documentation.

Table 4 Degrees of completeness of variables

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

Example: the variable “Daily wage/daily benefit” is only available for BeH and LeH observations; the observations of the other data sources contain the missing value “.n” for this variable. Another characteristic is that some variables have different contents depending on the data source. For instance, for BeH observations the “Employment status” variable contains the person group of the employment notification procedure, for LeH observations it contains the type of benefit, for LHG and XLHG observations it contains the SGB II status, for ASU and XASU observations the job search status and for MTH and XMTH observations it is the measure category. These differences are not immediately obvious from the variable name for every variable.

Table 5: List of variables with degree of completeness

List of variables	Page	BeH	LeH	LHG	ASU	XASU
Identifiers						
Individual ID (persnr)	38	h	h	h	h	h
IAB Establishment Panel ID (idnum)	38	h	l	l	l	l
IAB-Establishment-Panel-ID (idnum)	38	h	l	l	l	l
Observation counter per person (spell)	42	h	h	h	h	h
Source of spell (quelle)	43	h	h	h	h	h
Period of validity						
Original start date (begorig)	43	h	h	h	h	h
Original end date of observation (endorig)	44	h	h	h	h	h
Start date of split episode (begepi)	44	h	h	h	h	h
End date of split episode (endeipi)	44	h	h	h	h	h
Personal information						
Gender (frau)	45	h	h	h	h	h
Year of birth (gebjahr)	45	h	h	h	h	h
Month of birth (gebmon)	45	h	h	h	h	h
Nationality (nation)	45	h	h	h	h	h
Nationality, grouped (nation_gr)	46	h	h	h	h	h
Marital status (famst)	46	l	h	h	h	h
Number of children (kind)	46	l	h	h	m	m
Vocational training (ausbildung)	47	m	l	l	h	m
Vocational training (imputed) (ausbildung_imp)	50	h	l	l	l	l
School leaving qualification (schule)	50	m	l	l	h	m
Information on employment, benefit receipt and job search						
Daily wage, daily benefit rate (tentgelt)	51	h	h	l	l	l

List of variables	Page	BeH	LeH	LHG	ASU	XASU
Occupation - current/most recent (KldB 1988) (beruf)	52	h	l	l	m	l
Occupational group - current/most recent (KldB 2010) (beruf2010_3)	53	h	m	l	m	m
Occupational sub-group - current/most recent (KldB 2010) (beruf2010_4)	54	h	m	l	m	m
Level of requirement - current/most recent job (KldB 2010) (niveau)	55	h	m	l	m	m
Part-time (teilzeit)	55	h	l	l	l	l
Employment status (erwstat)	56	h	h	h	h	h
Transition zone (gleitz)	58	h	l	l	l	l
Temporary agency work (leih)	58	h	l	l	l	l
Fixed-term contract (befrist)	58	h	l	l	l	l
Reason of cancellation/notification/termination (grund)	58	h	h	m	h	m
Employment status prior to job search (estatvor)	60	l	l	l	h	m
Employment status after job search (estatnach)	60	l	l	l	m	l
Client profile (profil)	60	l	l	l	m	l
Type of termination of last job (art_kuend)	61	l	l	l	m	l
Desired working hours of the job sought (arbeitszeit)	61	l	l	l	m	l
Residual claim/planned duration (restanspruch)	61	l	h	l	l	l
Type of provider (traeger)	62	l	l	h	h	h
Start date of unemployment (alo_beg)	63	m	m	m	m	m
Duration of unemployment (alo_dau)	63	h	h	h	h	h
Location data						
Place of residence - district (Kreis) (wo_kreis)	63	h	h	h	h	h
Place of residence - federal state (Bundesland) (wo_bula)	64	h	h	h	h	h
Place of residence - employment agency (Arbeitsagentur) (wo_aa)	65	h	h	h	h	h
Place of residence - regional directorate (Regionaldirektion) (wo_rd)	65	h	h	h	h	h
Establishment variables						
Classification of economic activities 73 (w73_3)	66	h	l	l	l	l
Classification of economic activities 93, sub-classes (w93_5)	66	h	l	l	l	l
Classification of economic activities 93, groups (w93_3)	66	h	l	l	l	l
Classification of economic activities 03, sub-classes (w03_5)	67	h	l	l	l	l
Classification of economic activities 03, groups (w03_3)	67	h	l	l	l	l
Classification of economic activities 08, sub-classes (w08_5)	68	h	l	l	l	l
Classification of economic activities 08, groups (w08_3)	68	h	l	l	l	l
w73_3 completed by extrapolation/imputation (w73_3_gen)	68	h	l	l	l	l
Type of imputation w73_3 (group_w73_3)	69	h	l	l	l	l
w93_3 completed by extrapolation/imputation (w93_3_gen)	69	h	l	l	l	l
Type of imputation w93_3 (group_w93_3)	69	h	l	l	l	l

List of variables	Page	BeH	LeH	LHG	ASU	XASU
w08_3 completed by extrapolation/imputation (w08_3_gen)	70	h	l	l	l	l
Type of imputation w08_3 (group_w08_3)	70	h	l	l	l	l
Year of first appearance (grd_jahr)	70	h	l	l	l	l
First appearance (grd_dat)	71	h	l	l	l	l
Year of last appearance (lzt_jahr)	71	h	l	l	l	l
Last appearance (lzt_dat)	71	h	l	l	l	l
Total number of employees (az_ges)	72	h	l	l	l	l
Number of full-time employees (regular workers + others) (az_vz)	72	h	l	l	l	l
Number of employees in marginal part-time employment (az_gf)	72	h	l	l	l	l
Mean imputed wage all full-time employees (te_imp_mw)	73	h	l	l	l	l
Place of work - district (Kreis) (ao_kreis)	73	h	l	l	l	l
Place of work - federal state (Bundesland) (ao_bula)	74	h	l	l	l	l

1.6 Volume structure

The number of data rows in the administrative individual-level data is higher than the number of interviewed individuals stated here, because there are parallel information units on employment, receipt of benefits, and job seeking which each receive an individual data row. In addition, the data include all employees who are employed by the establishment at the time of the employer-survey. This also applies to the information on life, education, and employment history gathered during the interviews. Therefore, Table 6 indicates for how many individual process data is available, and how many rows they occupy in the administrative individual-level data. This is done separately for the total number of employees in at least waves and for all panel participants for whom there is a valid interview in both waves. Employees who have not given their consent to the merging of data are not included in the LPP-ADIAB 7517.

Table 6: Volume structure

Employees in cross section			
	Wave 1	Wave 2	Wave 3
Surveyed employees with linkage consent successfully linked with the IEB	6.576	6.460	6.180
Related Establishments from the LPP	825	709	715
Related lines from administrative personal data	347.294	334.001	315.055
Thereof: panel employees			
Surveyed employees with linkage consent successfully linked with the IEB		3.059	4.122
Related Establishments from the LPP		899	897
Related lines from administrative personal data		153.957	206.878

2 Data sources

2.1 LPP Establishment Survey

The establishment survey, which forms the first part of the LPP, was conducted as a follow-up survey of the IAB Establishment Panel in all waves. This has several benefits: on the one hand, the comprehensive data of the IAB Establishment Panel can be merged and used for detailed analyses. On the other hand, the LPP survey focuses on particularly relevant topics, because key parameters such as turnover or employee structure do not have to be surveyed again (Bellmann et al., 2014). The content of this additional survey includes “HR planning, recruitment, and development”, “Remuneration structure”, and “Loyalty, values, and corporate culture”. Moreover, structural attributes and changes in the establishment were enquired. The content of the survey was drawn up and developed together with the cooperation partners IAB, ZEW, and the Universities of Cologne

and Tübingen. The published LPP questionnaires and data reports⁵ provide a detailed overview of the interview topics covered in the establishment survey thus far.

The Kantar Public institute was responsible for conducting the establishment survey, including data review and weighting. The survey was carried out verbally and in person, analogous to the IAB Establishment Panel, with the option of leaving the questionnaire for self-completion in the establishment. The respondent was the person in the company who also answered the questionnaire on the IAB Establishment Panel (Tschersich and Gensicke 2015). For more information on the establishment survey, please refer to the LPP data and method reports published on the FDZ website.

2.2 LPP Employee Survey

The employee survey forms the second part of the LPP and is always conducted a few months after the establishment survey. It covers the employees' point of view and adds individual data on the following key areas to the establishment data: "personal characteristics and employment", "personnel development", "working conditions and burdens", "remuneration", "commitment, values and corporate culture", "personality and attitudes", "health" and "sociodemography". The published LPP questionnaires and data reports⁶ provide a detailed overview of the interview topics covered in the employee survey thus far.

In addition to linking establishment and employee data, the dataset also offers comprehensive employee-level analyses. The content of the survey was also drawn up and developed by the cooperation partners listed above. The infas Institute for Applied Social Sciences was tasked with the implementation of the LLP employee survey and the documentation of the field study phase. The survey was conducted as a telephone survey (CATI) and pre-tested in advance. For more information on the employee survey, please refer to the data and method reports, which are freely available on the FDZ website.

2.3 Survey Data from the IAB Establishment Panel

The IAB Establishment Panel is an annual representative establishment survey on establishment employment characteristics on the reference date of 30 June. Apart from core and exchangeable modules, the catalogue of questions also includes one-time focus modules. The IAB Establishment Panel includes thousands of establishments of any size from all walks of industry every year. As a general rule, the survey is conducted in the form of oral one-on-one interviews. Respondents may, however, choose to complete the questionnaire by themselves and then send it to Kantar, just like they can for the Establishment Panel. In parts of Saarland and Schleswig-Holstein, the survey was conducted via post. For more information on the IAB Establishment Panel, please refer to the FDZ website.

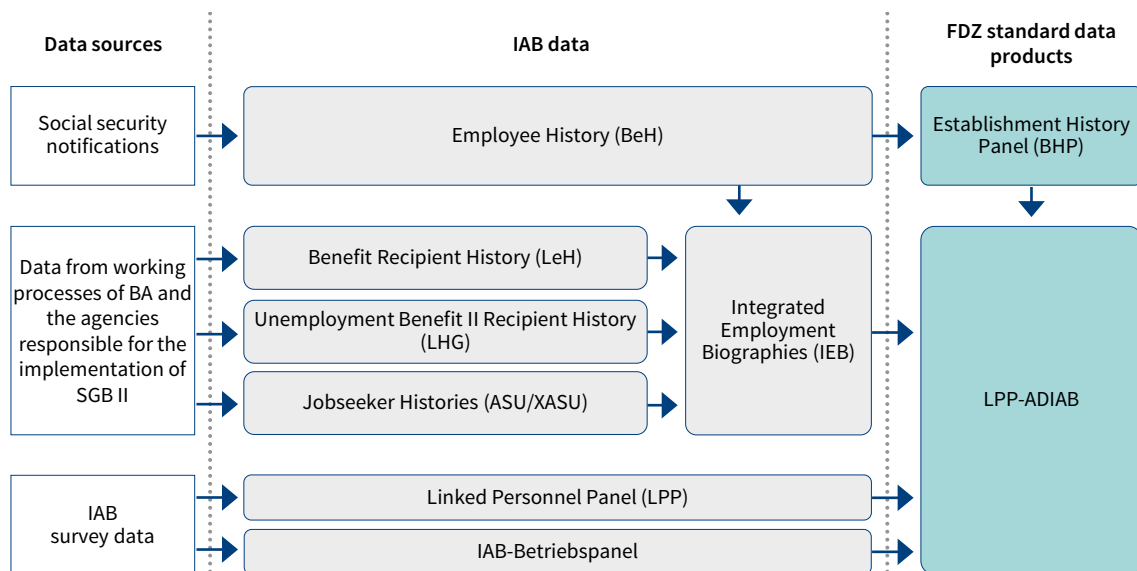
⁵ LPP questionnaires and data reports: https://fdz.iab.de/en/Integrated_Establishment_and_Individual_Data/lpp/LPP1617.aspx

⁶ LPP questionnaires and data reports: https://fdz.iab.de/en/Integrated_Establishment_and_Individual_Data/lpp/LPP1617.aspx

2.4 Administrative Data from the IEB

The administrative individual data were drawn from the Integrated Employment Biographies (IEB) of the IAB. They unite data from different data sources, each of which may contain information from different administrative procedures. Figure 2 illustrates the data streams leading to the LPP-ADIAB 7517.

Figure 2: Data sources of the LPP-ADIAB 7517



2.4.1 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 1 January 1973 (and was extended to cover East Germany as of 1 January 1991) and is known by the abbreviation DEÜV (previously DEVO / DÜVO) (for further details see: Bender et al. 1996, p. 4 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. 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 students⁷ (see Cramer 1985) are not recorded in the BeH in principle. Since the notification procedure was changed on 1 January 1999, employees in marginal part-time employment and unpaid family workers have also been recorded (not contained in the data until 1 April 1999). The data are recorded by the health insurance companies, collected and edited by the Federal Employment Agency (BA) and subsequently integrated into the History File by the IAB.

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

⁷ 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, one has to take into account that the variables in both the Basis Establishment File and the BHP Extension Files are aggregated on 30 June of a year. Unlike the data on individuals, the establishment variables are therefore not spell data but are only valid on 30 June precisely each year.⁸

2.4.2 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. Since 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 Benefit Recipient History. The earliest available data in the LeH are from 1 January 1975.

2.4.3 Unemployment Benefit II Recipient History (LHG)

The Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung - LHG) contains information about individuals who are eligible for benefits and capable of work, about the members of their benefit unit (Bedarfsgemeinschaft) in accordance with § 7 SGB II and about certain individuals associated with the benefit unit. In the SIAB, however, it is not possible to link individuals with benefit receipt in accordance with Social Code Book II (SGB II) within benefit units. The receipt of benefits in accordance with SGB II covers both basic social security benefits (e.g., Unemployment Benefit II) and supplements to unemployment benefit or additional benefits. The LHG does not contain any information about the benefit rates, however. As the amount of benefit received is not determined at the level of the individual but at the level of the benefit unit in the case of Unemployment Benefit II, it is difficult to assign an individual benefit rate. Therefore, the LPP-ADIAB also does not contain information about SGB-II-benefit rates.

Unlike the benefits in the sphere of Social Code Book III, the Federal Employment Agency (BA) is not the sole institution responsible for administering the benefits. The data therefore distinguish between the three possible types of institution responsible for implementing SGB II:

- Cooperation of employment agencies and municipalities (Arbeitsgemeinschaften – ARGE) until the end of 2010 / joint facilities (gemeinsame Einrichtungen) since 2011), in which the BA and the municipality deal with tasks jointly,
- separated responsibilities (getrennte Trägerschaft - gt) / municipalities exercising their duties separately (until 2011) – here the tasks are divided between the BA and the municipality,⁹

⁸ An extreme example: an employment notification exists from 1 January 2006 to 30 May 2006; the establishment goes bankrupt in June 2006. In this case, there would be no information about this establishment in the BHP for 2006.

⁹ The municipality pays the costs for housing and heating (Section 22 SGB II) and additional one-off benefit payments to cover extra costs (Section 23 (3) SGB II) and the additional benefits to support integration in accordance with Section 16 (2) Clause 2 No. 1 - 4 SGB II. The BA, on the other hand, covers the costs for regular benefits, social security contributions and integration benefits (SGB III and SGB II) and specific benefits excluding the additional benefits to support integration cited above.

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

The data of the “Unemployment Benefit II Recipient History drawn from A2LL” (LHG) come from different reporting procedures. As a rule, the IT procedure A2LL was used in all ARGE cooperation projects until 2010, and in joint facilities from 2011 onwards.¹⁰ Authorised municipalities use various IT procedures of their own and transmit their data to the BA by means of the XSozial-BA-SGB II standard once a month. Both of the procedures are used by municipalities with separated responsibilities. The different data standards affect the scope and quality of the data supplied.

The earliest available data in the LHG are from 1 January 2005. However, the data source is incomplete until the beginning of 2007 (see Section 4).

2.4.4 Jobseeker Histories (ASU/XASU)

Data about jobseekers are stored in the Jobseeker Histories (Arbeitsuchendenhistoriken – ASU/XASU). The ASU data source contains information on jobseekers who are registered with employment agencies, and from 2005 onwards also includes ARGE cooperation projects and separated responsibilities for the implementation of SGB II. The XASU data source, on the other hand, contains the data of jobseekers in receipt of Unemployment Benefit II (ALG-II) from authorised municipalities from 2005 onwards. These data are reported in accordance with the X-Sozial-BA-SGB II standard. The earliest available data in the ASU are from 1 January 1997 and in the XASU from 1 January 2005, respectively.

3 Data preparation

3.1 LPP survey data

3.1.1 Dataset descriptions

The sampling procedure and data-collection, -structure, -preparation and -editing within the scope of the LPP Employer Survey study are described in the method reports 03/2015 (Tschersich and Gensicke 2015), 03/2016 (Tschersich and Gensicke 2016) und 12/2018 (Tschersich and Gensicke 2016). The same information for the employee survey is available in the method reports 02/2015 (Schütz et al. 2015), 02/2016 (Schütz et al. 2016) und 11/2018 (Schütz et al. 2017). The questions used in the interviews are fully reflected in the FDZ data reports 01/2015 (Broszeit and Wolter 2015), 06/2016 (Broszeit et al. 2016) and 04/2018 (Mackeben et al. 2018). The questionnaires can be downloaded from https://fdz.iab.de/de/Integrated_Establishment_and_Individual_Data/lpp/lpp-adiab.aspx.

3.1.2 Data preparation changes

The LPP-ADIAB version of the LPP employee survey only contains people who have consented to have their data linked to the administrative data. During the clean-up procedure it emerged that

¹⁰ In 2014, A2LL was gradually replaced by ALLEGRO as the new IT procedure for Unemployment Benefit II in the sphere of SGB II in joint facilities.

two people had participated in the second and third waves, respectively, under different artificial individual IDs. In order to facilitate a linking with administrative data, these cases were merged in a unified ID (persnr) in the employee survey data. Moreover, it was found that three persons had participated in the survey twice during the third wave. During data preparation, their entries were kept under the original artificial individual ID; the duplicate entries were removed. This ensures that the panel character of the dataset and the possibility to link it with the administrative data are preserved. Missing values were allocated to the betnr and idnum of individuals who are not employed in the same establishment as in the previous wave. The assignment to an establishment from the IEB is left to the user in these cases.

3.2 Administrative data from the IEB

3.2.1 Corrections and validation procedures

Before the data from the sources specified in Section 2 are merged to form the IEB, they undergo source-specific correction procedures (see the following Sections). The IEB as a whole undergoes the following corrections which are analogous to SIAS 7517 (Antoni et al. 2019):

- 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.
- Records with no information on the date of birth or on gender after the correction procedure are deleted.

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

3.2.2 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, and 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 31 December 2017.

3.2.3 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, since 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 the BeH in the LPP-ADIAB.

3.2.4 Unemployment Benefit II Recipient History (LHG)

- Observations without a BA client number are deleted.
- Cancelled data records are not used.
- It only contains observations of people who are capable of work and people under the age of 65.
- In each case, non-overlapping periods of benefit entitlement of a person in a certain benefit unit (BG) are created. New observations are begun for the following administrative reasons:
 - on certain birthdays of members of the BG that are stipulated by law and relevant for structural changes in the benefit unit (14, 15, 18 and 25) and the individual retirement age of members of the BG (see Section 3.2.6),
 - when the structure of the benefit unit changes (e.g. due to entries/exits),
 - when there are changes in a variable of the BG client and
 - at the beginning and the end of a case of benefit sanctions for observations from 1 April 2006 onwards. It must be taken into account, however, that it is not possible to identify the duration or type of sanction or the time when it was imposed or when it began on the basis of the data. The reason for this is that there is no corresponding variable or value that indicates the start, type or duration of the sanction.
- For the reason mentioned above, all individual-related variables that are available for the LHG source are valid for the entire duration of the observation.
- Double notifications due to the territorial reforms in 2009/2011 and the reorganisation of the institutions in 2012 (see Section 4.1.2) were corrected as far as possible.
- The territorial allocations are corrected in the same way as the BeH in the LPP-ADIAB.

3.2.5 Jobseeker Histories (ASU/XASU)

- Observations with an end date before 1 January 1995 are not included.

- There is no consolidation of the ASU observations for individual persons. Therefore, overlaps between ASU observations might occur.
- Individual-related variables that are only available for the (X)ASU sources always refer to the beginning of the spell.
- A new ASU spell is generated as soon as a change of status occurs (e.g. from seeking work to unemployed). This also applies if the type of institution (employment agency, cooperation of employment agency and municipality, joint facility, authorised municipalities, separated responsibilities) changes.
- The XASU contains non-overlapping time periods for individuals. If one of the following variables changes, in each case a new data spell is generated for the XASU:
 - change of job search status
 - change of availability
 - change of SGB II institution (due to notification procedure)
 - change of place of residence
- The territorial allocations are corrected in the same way as the BeH in the LPP-ADIAB.

3.2.6 SGB II anonymization

In order to reduce the risk of de-anonymisation, in general only the year of birth is indicated in the administrative data. The month of birth can only be requested as a sensitive variable if there is a justified need for it. However, in the LHG and (X)ASU there is the risk that the exact date of birth may still be obvious due to the chronological structure of the observations. Observations might end systematically on certain birthdays and/or the day before, or start again on the birthday.

In order to prevent an indirect determination of the exact date of birth, the following procedure is applied. Observations split on the 18th, 25th or 65th birthday are merged into a single observation. Apart from the exact date of birth, no other information is lost in this case. For observations beginning on the 15th birthday or on the first day of retirement (or ending on the first day of retirement), the start date (end date) of the observation is set to the beginning (the end) of the respective quarter. This correction results in a bias of the duration of the observation. Corrected observations are marked in the variable "Employment status".

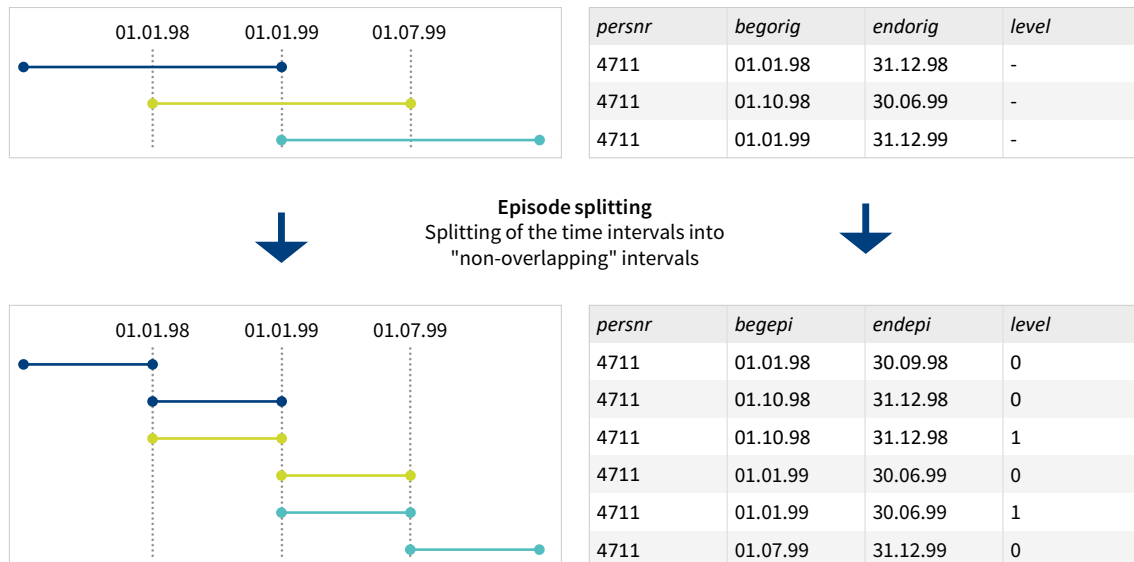
3.2.7 Episode splitting

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 3).

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

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

Figure 3: Episode splitting



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 LPP-ADIAB can be generated using the following Stata commands if required (see Box 5).

Box 5: Example code to create additional observation counters in Stata 14

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

3.2.8 Sampling procedure

The IEB excerpt contains the administrative data of all employees in the LPP Employee Survey, provided they have consented to this linking. In addition, it includes the administrative data of all persons who were employed by the establishments in the LPP Establishment Survey at the time of the survey. Moreover, the BHP includes all establishments that employed employees from the LPP Employee Survey who consented to have their data merged.

3.2.9 Missing values

Table 7: Missing values

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 (dark grey cells in the overview of variables in Section 1.5) or is not available for a certain period.

4 Data quality and problems

4.1 Integrated Employment Biographies (IEB) in General

4.1.1 Gaps in employment histories

The IEB contains 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 ‘reason for notification / reason for end of benefit receipt / reason for discontinuation of SGB II / reason for deregistration’ variable of the observation immediately preceding the gap (if a corresponding observation exists).

These gaps were identified using the variables “Reason of cancellation/notification/termination” and “Employment status” in the various sources. The list in Table 8 makes no claims to be exhaustive.

Table 8: 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 3.2.5)	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 SGB 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 1 January 2005, the responsibilities for the care of job seekers 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 has also been 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 job seekers until 31.12.2010. With a constitutional amendment, the option was extended to 110 providers from 2012 onwards.

While the ARGE/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 1 January of 2011 and of 2014, various ARGE/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 2.4.1) using a so-called occupation code. The new occupation code 2010 was adopted for notifications with an end date later than 30 November 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 level. 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.¹¹

¹¹ See Paulus and Matthes, 2013, for details regarding the Classification of Occupations 2010.

- 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,¹² 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 and Thomsen (2016) developed an imputation procedure to replace the missing values by imputed values. The imputed data are included in the LPP-ADIAB 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 1 April 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-month 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-month file shows a 100% degree of completeness by definition.
- In the version of the IEB the, the year 2014 is the last year with a degree of completeness of BeH observations of 100%.¹³ For the years 2015 and 2016, the 18-month files were used, and the observations for 2017 originate from a 6-month file. It can therefore be assumed that employment notifications for 2015 and 2016 are slightly underreported in the SIAB, 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,

¹² The test programs used in the notification procedure permitted missing details in the occupation code 2010 until the end of May 2012.

¹³ 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-month files were available instead of the 36-month files. However, analyses with earlier data versions have shown that, as a rule, the 18-month file already has a degree of completeness of around 99%. This means that in the next 18 months there will only be extremely few follow-up or correction notifications.

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 Hoyer-swerda (14264).
- Concerning the notifications for full-time employment, especially the districts Main-Taunus (06436) and Alzey-Worms (07331) are noteworthy. They feature above-average rises. Also in this case, the reasons are notification problems 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' 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 company relocations between these regions, the following adjustment rule was implemented at the level of the company number: As soon as the company 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 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 1 July 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.

4.4 Unemployment Benefit II Recipient History (LHG)

- With regard to the completeness of case numbers or benefit histories from the LHG data sources, there are substantial gaps in the years 2005 and 2006. We therefore strongly advise against analysing the data for this time period based merely on the LHG sources.
- Longitudinal analyses of individuals are affected by inaccuracies as it is not possible to distinguish between changes in the benefit entitlement status and relocations into and out of districts whose institutions had problems delivering data.
- Also from 2007 onwards, cases of underrecording occur at times. These generally last one month and occur mainly in the authorised municipalities.
- Underrecording and overrecording occur in connection with changes in the type of institution responsible for implementing SGB II:
 - In the context of the reform of the territories covered by the institutions, which came into force on 1 January 2011, cases of underreporting occurred in the districts covered by the employment agencies of Dessau-Roßlau, Halberstadt, Halle and Sangerhausen.
 - Double notifications due to the territorial reforms in 2009/2011 and the changes in the form of the institutions as of 1 January 2012 and of 1 January 2013 are already corrected as far as possible in the IEB. Nonetheless double notifications may still occur.
- In the following job centres there are inaccuracies with regard to the allocation of benefit cases:
 - between Emden and Norden between September and December 2009
 - between Döbeln and Mittelsachsen from October to December 2012
 - between Tirschenreuth and Wunsiedel from November 2012 to March 2013
- Some individuals for whom a (X)LHG spell exists are excluded entirely or partly from benefit receipt according to SGB II, for instance because they take part in a subsidised training programme, receive an old-age pension, live in an in-patient facility or a residential institution or receive insurance payments aimed at avoiding need. This affects on average 3 to 5 percent of all cases. In XSozial this person group is sometimes underrecorded by some institutions. Exclusion from benefits cannot be identified in the LPP-ADIAB 7517.
- Due to the reporting logic, information from the XSozial transmission standard can only be updated monthly.
- In the official performance statistics of the BA, reporting gaps are supplemented by a statistical estimation procedure at an aggregated regional level. No supplementary data sets are provided in the LHG. A comparison of the IEB with the performance statistics is therefore only possible to a limited extent.

4.5 Jobseeker Histories (ASU/XASU)

4.5.1 ASU

- The registered periods of job search activity in the ASU source are regarded as complete from the year 1997 onwards. Therefore, the analysis potential of the ASU spells before 1997 is limited.
- With the introduction of SGB II on 1 January 2005, jobseekers are no longer fully covered by BA procedures. From this date, the ASU only covers persons who are supported by the BA in the sphere of the SGB III (employment promotion) or by ARGE, gE or gT in the sphere of the SGB II (basic security).
- For the placement staff it is not always possible to record the allocation to the legal sphere immediately, since it is frequently only clear which institution is primarily responsible after a certain time due to a possible entitlement to SGB II benefits. Therefore, we recommend comparing the value of the variable "Type of provider" in the ASU with the value in the LHG and/or XLHG for the same period of time. Due to the recording gaps in the LHG and XLHG between 2005 and 2006 this is not always possible.
- For some individuals for whom an authorised municipality has been responsible since 2005, parallel "artificial" ASU datasets were created by the Federal Employment Agency. These can be identified via *estatvor* (transfer to an authorised municipality).
- From mid-2005 until mid-2006, the coArb IT procedure, from which the jobseeker and applicant pool data originate, was superseded by the VerBIS procedure at the Federal Employment Agency. In July 2005, coArb was first replaced by VerBIS in the employment agency in Wiesbaden as a pilot project. From December 2005 onwards, it was then gradually replaced by VerBIS in several stages in all employment agencies. The information for many of the variables recorded was gathered with different levels of differentiation and different qualitative weighting in the two systems. It is therefore very difficult to integrate these variables into the IEB, which is only possible using a special procedure (mapping). Unfortunately, a full conversion of the affected variables from coArb to VerBIS cannot be achieved by means of mapping, so for some variables there is an unusually large number of the values 'no details available', 'other' or 'missing'. Moreover, striking differences may occur in frequency counts, depending on whether the original source of the data was coArb or VerBIS. It can be assumed that integration agreements are under-recorded. Important limitations in the analysis potential are mentioned in the corresponding description of variables.
- The coArb procedure, which was used until June 2006, supported only the placement of unemployed persons and jobseekers. Some data were also collected about individuals who were only seeking advice, but these data are incomplete. The careers advice data were collected in a separate system. In VerBIS the attributes of the job-search status were extended to include 'seeking advice' and individuals 'without status'. The latter group includes individuals eligible for Unemployment Benefit II who are only available for job placement to a limited degree. The recording of this group in VerBIS is only regarded as largely complete since January 2008.

- A change of the institution responsible for implementing SGB II or a change of place of residence does not lead to a new ASU observation, the value of the variable at the start of an episode is continued. The longer the observation becomes, the greater the risk is that the institution responsible or the place of residence is no longer correct.
- The job search status is hardly corrected afterwards, which is why the case numbers largely coincide with the BA statistics until 2005. Since spring 2011, the jobseeker data from BA procedures and XSozial have been consolidated in the integrated unemployment statistics. This may lead to larger deviations. At the current margin, however, the stock data are identical.

4.5.2 XASU

- In contrast to the job search spells from the cooperation of employment agencies and municipalities (ARGE) and the separated responsibilities, systematic cases of underrecording have emerged for the authorised municipalities since 1 January 2005. Thus, data from the XASU source should only be analysed from 2007 onwards.
- From 2007 onwards, individual months are also repeatedly affected by delivery failures, analogous to the benefit receipt data (LHG).
- Over-reporting may also occur at certain points. A large number of technical problems can be responsible for this, including systematically missing deregistrations, incorrect reversals of cancellations and parallel job search notifications in several districts.
- A variety of variables sometimes have only a very low degree of completeness for the XASU. Variables which are affected by this include 'School leaving qualification', 'Reason of notification' as well as 'Employment status prior to job search'. Although the degree of completeness of these variables improves over time, some of them are still unsatisfactory. The 'Occupation - current/most recent' variable is not available in the XASU for almost the entire period available.
- For a number of institutions (districts), the proportion of registered recipients of unemployment benefit II who are also registered jobseekers is implausibly large at times or continuously in the IEB. One possible reason for this could be an incorrect determination of the status 'not unemployed but seeking work' by these institutions.
- The institution-related and period-related plausibility of the XASU data should be examined before use, taking the research question into account.
- Due to the reporting logic, information from the XSozial reporting procedure can only be updated monthly.
- Differences in consolidation rules, time references and regional assignments may result in differences to the published BA statistics.

5 Description of variables

Frequency counts and overviews of the individual values and labels of the variables can be found in separate files under <http://fdz.iab.de/en.aspx>. This chapter lists variables of the administrative data. For an overview on the survey variables, please refer to the data reports of the LPP.

5.1 Identifiers

5.1.1 Individual ID (persnr)

Variable label	Individual ID
Variable name	persnr
Category	identifiers
Origin	BeH, LeH, LHG, ASU, XASU
Data type	numerical
Detailed description	<p>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.</p> <p>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.</p> <p>The formation of the individual identifier which spans all data sources is based on a heuristic developed by the BA.</p>

5.1.2 IAB-Establishment-Panel-ID (idnum)

Variable label	IAB-Establishment-Panel-ID
Variable name	idnum
Category	identifiers
Origin	LPP Employer Survey, IAB Establishment Panel
Data type	numerical
Detailed description	The IAB-Establishment-Panel-ID can be used for the linkage between LPP Employer Survey and IAB Establishment Panel.

5.1.3 Establishment ID (betnr)

Variable label	Establishment ID
Variable name	betnr
Category	identifiers
Origin	IEB
Data type	numerical
Detailed description	<p>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.) This identifier only allows to links individuals and establishments within IEB data.</p>

	<p>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.</p> <p>For the establishment number, the following should be observed in general:</p> <p>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.</p> <p>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.</p> <p>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.</p> <p>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:</p> <p>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.</p> <p>A workplace is a unit in which employees work and which is not allocated an establishment number according to the above-mentioned principles.</p> <p>A company as a term combines establishment premises and workplaces belonging to the same employer.</p> <p>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.</p> <p>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.</p>
Notes on quality	<p>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.</p>

5.2 Variables from survey para data

5.2.1 Interview date (intdat)

Variable label	Date of Interview
Variable name	intdat
Category	Para data from the survey
Origin	LPP employee survey
Data type	numerical

Detailed description	The interview date reflects the exact date of the LPP employee survey. It is coded in DDMMYYYY.
----------------------	---

5.2.2 Weighting factor LPP Employee Survey (wave 1)

Variable label	Weighting factor LPP Employee Survey
Variable name	gew_pers_2012_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2012/2013
Data type	numerical
Detailed description	Adjusted weighting factors. The adjustment is described in section 1.3.

5.2.3 Weighting factor LPP Employee Survey: Cross (wave 2)

Variable label	Weighting factor LPP Employee Survey: Cross
Variable name	gew_pers_2014_quer_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2015
Data type	numerical
Detailed description	Adjusted cross-sectional weight valid for all panel cases and refreshers Employment group: „Without movers and without exits“ The adjustment is described in section 1.3.

5.2.4 Weighting factor LPP Employee Survey: Cross (mover, wave 2)

Variable label	Weighting factor LPP Employee Survey: Cross (mover)
Variable name	gew_pers_2014_quer_m_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2015
Data type	numerical
Detailed description	Adjusted cross-sectional weight valid for all panel cases and refreshers Employment group: „With movers and without exits“ The adjustment is described in section 1.3.

5.2.5 Weighting factor LPP Employee Survey: Cross (mover, exits, wave 2)

Variable label	Weighting factor LPP Employee Survey: Cross (mover, exists)
Variable name	gew_pers_2014_quer_m_e_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2015
Data type	numerical
Detailed description	Adjusted cross-sectional weight valid for all panel cases and refreshers Employment group: „With movers and with exits“ The adjustment is described in section 1.3.

5.2.6 Weighting factor LPP Employee Survey: Panel (wave 2)

Variable label	Weighting factor LPP Employee Survey: Panel
Variable name	gew_pers_2014_panel_ADIAB

Category	Weighting factors
Origin	LPP employee survey 2015
Data type	numerical
Detailed description	Adjusted panel weight valid for all panel cases Employment group: „Without movers and without exits“. The adjustment is described in section 1.3.

5.2.7 Weighting factor LPP Employee Survey: Panel (mover, wave 2)

Variable label	Weighting factor LPP Employee Survey: Panel (mover)
Variable name	gew_pers_2014_panel_m_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2015
Data type	numerical
Detailed description	Adjusted panel weight valid for all panel cases Employment group: „With movers and without exits“. The adjustment is described in section 1.3.

5.2.8 Weighting factor LPP Employee Survey: Panel (mover, exits wave 2)

Variable label	Weighting factor LPP Employee Survey: Panel (mover, exits)
Variable name	gew_pers_2014_panel_m_e_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2015
Data type	numerical
Detailed description	Adjusted panel weight valid for all panel cases Employment group: „With movers and with exits“. The adjustment is described in section 1.3.

5.2.9 Weighting factor LPP Employee Survey: Cross (wave 3)

Variable label	Weighting factor LPP Employee Survey: Cross
Variable name	gew_pers_2016_quer_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2017
Data type	numerical
Detailed description	Adjusted cross-sectional weight valid for all panel cases and refreshers Employment group: „Without movers and without exits“ The adjustment is described in section 1.3.

5.2.10 Weighting factor LPP Employee Survey: Cross (mover, wave 3)

Variable label	Weighting factor LPP Employee Survey: Cross (mover)
Variable name	gew_pers_2016_quer_m_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2017
Data type	numerical
Detailed description	Adjusted cross-sectional weight valid for all panel cases and refreshers Employment group: „With movers and without exits“ The adjustment is described in section 1.3.

5.2.11 Weighting factor LPP Employee Survey: Cross (mover, exits, wave 3)

Variable label	Weighting factor LPP Employee Survey: Cross (mover, exists)
Variable name	gew_pers_2016_quer_m_e_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2017
Data type	numerical
Detailed description	Adjusted cross-sectional weight valid for all panel cases and refreshers Employment group: „With movers and with exits“ The adjustment is described in section 1.3.

5.2.12 Weighting factor LPP Employee Survey: Panel (wave 3)

Variable label	Weighting factor LPP Employee Survey: Panel
Variable name	gew_pers_2016_panel_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2017
Data type	numerical
Detailed description	Adjusted panel weight valid for all panel cases Employment group: „Without movers and without exits“. The adjustment is described in section 1.3.

5.2.13 Weighting factor LPP Employee Survey: Panel (mover, wave 3)

Variable label	Weighting factor LPP Employee Survey: Panel (mover)
Variable name	gew_pers_2016_panel_m_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2017
Data type	numerical
Detailed description	Adjusted panel weight valid for all panel cases Employment group: „With movers and without exits“. The adjustment is described in section 1.3.

5.2.14 Weighting factor LPP Employee Survey: Panel (mover, exits wave 3)

Variable label	Weighting factor LPP Employee Survey: Panel (mover, exits)
Variable name	gew_pers_2016_panel_m_e_ADIAB
Category	Weighting factors
Origin	LPP employee survey 2017
Data type	numerical
Detailed description	Adjusted panel weight valid for all panel cases Employment group: „With movers and with exits“. The adjustment is described in section 1.3.

5.3 Generated technical variables

5.3.1 Observation counter per person (spell)

Variable label	Counter per person
Variable name	spell
Category	generated technical variables
Origin	BeH, LeH, LHG, ASU, XASU

Data type	numerical
Detailed description	The observation counter per person counts a person's observations, beginning with 1. The variable is generated during the episode splitting procedure and refers to the split observations. Using the "observation counter per person" variable, it is easy to restore the original sorting order. The observations are sorted first by the start date of the split episode and then by the data source.

5.3.2 Source of spell (quelle)

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

5.3.3 Year (jahr)

Variable label	Year
Variable name	jahr
Category	generated technical variables
Origin	BeH
Data type	numerical
Detailed description	This variable is only included in the Basis Establishment File. It indicates the year of validity of the establishment data as of the reference date of 30 June. This variable can be used together with the establishment number to link the Individual File and the Basis Establishment File. See Box 1 for an example code with the „merge“-command in Stata 14.

5.4 Period of validity

5.4.1 Original start date (begorig)

Variable label	Original start date
Variable name	begorig
Category	period of validity
Origin	BeH, LeH, LHG, ASU, XASU
Data type	date
Detailed description	<p>The original start date of the observation corresponds to the original start date of the notification. This can differ from the start date of the episodes (begepi) (see also the comments on episode splitting in section 3.2.7)</p> <p>Since the notification logic might under certain circumstances permit re-identification of the exact day of birth, the original information on the date was changed in these cases by applying the anonymisation procedure described in Section 3.2.6.</p> <p>1) BeH</p> <p>Because of the rules of the notification procedure, in BeH observations the starting and ending year are always identical (obligation of the employer to submit annual employment notifications). A continuous employment relationship may therefore be distributed across several notifications.</p> <p>2) LHG, ASU, XASU</p>

	Certain changes lead to the creation of a new observation (see Sections 3.2.4 and 3.2.5). begorig indicates the start date of the new period.
--	---

5.4.2 Original end date of observation (endorig)

Variable label	Original end date
Variable name	endorig
Category	period of validity
Origin	BeH, LeH, LHG, ASU, XASU
Data type	date
Detailed description	<p>The original end date of the observation corresponds to the original end date of the notification. This can differ from the end date of the relevant line of data, the so-called end date of the split episode (see also the comments on episode splitting in Section 3.2.7).</p> <p>Since the notification logic might under certain circumstances permit re-identification of the exact day of birth, the original information on the date was changed in these cases by applying the anonymisation procedure described in Section 3.2.6</p> <p>1) BeH 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.</p> <p>2) LHG, ASU, XASU Certain changes lead to the creation of a new observation (see Sections 3.2.4 and 3.2.5) endorig indicates the end date of the new period.</p>

5.4.3 Start date of split episode (begepi)

Variable label	Episode start date
Variable name	begepi
Category	generated period of validity
Origin	BeH, LeH, LHG, ASU, XASU
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 3.2.7).

5.4.4 End date of split episode (endepe)

Variable label	Episode end date
Variable name	endepe
Category	generated period of validity
Origin	BeH, LeH, LHG, ASU, XASU
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 3.2.7).

5.5 Personal information

5.5.1 Gender (frau)

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

5.5.2 Year of birth (gebjahr)

Variable label	Year of birth
Variable name	gebjahr
Category	personal variables
Origin	BeH, LeH, LHG, ASU, XASU
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.

5.5.3 Month of birth (gebmon)

Variable label	Month of birth
Variable name	gebmon
Category	personal variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	<p>The month of birth is constant within one individual account.</p> <p>One can use the variables Year of birth (gebjahr) and Month of birth (gebmon) to generate a variable in the date format JJJmM (e.g., 1984m6) with the following syntax in Stata:</p> <pre>gen int gebdat = ym(gebjahr, gebmon) format gebdat %tm</pre>
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 (gebjahr) is provided.

5.5.4 Nationality (nation)

Variable label	Nationality
Variable name	nation
Category	personal variables
Origin	BeH, LeH, LHG, ASU, XASU
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 (nation_gr) is provided.

5.5.5 Nationality, grouped (nation_gr)

Variable label	Nationality, grouped
Variable name	nation_gr
Category	personal variables
Origin	BeH, LeH, LHG, ASU, XASU
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.

5.5.6 Marital status (famst)

Variable label	Marital status
Variable name	famst
Category	personal variables
Origin	LeH, LHG, ASU, XASU
Data type	numerical
Detailed description	<p>This variable describes the marital status.</p> <p>1) LeH In the LeH, the variable has only two values (0 - not married, 1 - married).</p> <p>2) LHG, ASU, XASU In the sources LHG, ASU and XASU, a distinction is made between six values (values 11-16). The information from the different sources was not compared.</p>
Notes on quality	<p>1) LeH 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%.</p> <p>2) LHG, XASU Due to deviating reporting standards in XSozial, the information from the zkT has to be classified as less valid until December 2009.</p>

5.5.7 Number of children (kind)

Variable label	Number of children
Variable name	kind
Category	personal variables

Origin	LeH, LHG, ASU, XASU
Data type	numerical
Detailed description	<p>This variable has a different meaning depending on the data source.</p> <p>1) LeH 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: 0 no children 100 one or more children</p> <p>2) ASU In these sources, the value of the variable corresponds to the actual number of children under 15 living in the household.</p> <p>3) LHG, XASU The variable reports the actual number of children aged under 15 in the benefit unit (Bedarfsgemeinschaft). In the LHG, the value is valid for the entire original period.</p>
Notes on quality	<p>1) 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.</p> <p>3) ASU Until 30 June 2006, only up to nine children could be recorded. The value zero does not exist. For observations prior to 30 June 2006, the value zero was recoded to "missing", since it is not clear whether zero should be interpreted as "no children" or as "field not filled in". For observations after 30 June 2006, the variable is only recorded if children exist.</p>

5.5.8 Vocational training (ausbildung)

Variable label	Vocational training
Variable name	ausbildung
Category	personal variables
Origin	BeH, ASU, XASU
Data type	numerical
Detailed description	<p>It must be taken into account that this variable has a different meaning depending on the data source:</p> <p>1) BeH 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: 1 Without vocational training 2 In-company voc. training/traineeship/external voc. training 11 University of applied sciences without further specifications</p>

	<p>12 University without further specifications</p> <p>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.</p> <p>2) ASU</p> <p>For these observations the vocational education completed most recently is reported. The following values exist in spells with a start date until 30 June 2006:</p> <p>1 Without vocational training</p> <p>2 In-company voc. training/traineeship/external voc. training</p> <p>3 Technical school (voc. training)</p> <p>4 Technical school (advanced voc. training)</p> <p>5 University of applied sciences (FH)</p> <p>6 University</p> <p>In 2006, the IT procedure from which the jobseeker data originate was switched from coArb to VerBIS. Many variables, such as training, were reported with different levels of differentiation in the two systems. This means that in spells from ASU which have a start date from 1 July 2006 onwards, the following differentiated categories are available:</p> <p>7 Voc. training not accepted in Germany</p> <p>8 University degree not accepted in Germany</p> <p>9 In-company voc. training/traineeship/in-school voc. training</p> <p>10 Other exams</p> <p>11 University of applied sciences without further specifications</p> <p>12 University without further specifications</p> <p>13 Doctorate</p> <p>14 Bachelor (BA)</p> <p>15 Bachelor (FH)</p> <p>16 Bachelor (University)</p> <p>17 Master (FH)</p> <p>18 Master (University)</p> <p>19 Diploma (BA)</p> <p>20 Diploma (FH)</p> <p>21 Diploma (University)</p> <p>23 Undergraduate studies</p> <p>24 Secondary/additional studies</p> <p>25 Other (promotion-) advanced training for graduates</p> <p>3) XASU</p> <p>For spells that originate from these sources, the vocational education completed most recently is reported. The following values exist:</p> <p>1 Without vocational training</p> <p>2 In-company voc. training/traineeship/external voc. training</p>
--	---

	<div>3 Technical school (voc. training)</div> <div>4 Technical school (advanced voc. training)</div> <div>5 University of applied sciences (FH)</div> <div>6 University</div> <div>7 Voc. training not accepted in Germany</div> <div>8 University degree not accepted in Germany</div> <div>For analyses that cover a longer period of time, the values can be aggregated as follows:</div> <table><tr><th>Ausbildung</th><th>ausbildung_agg</th><th>Value label</th></tr><tr><td>1, 22</td><td>1</td><td>Without (recognised) vocational training</td></tr><tr><td>7</td><td>2</td><td>Vocational training not accepted in Germany</td></tr><tr><td>8</td><td>3</td><td>University degree not accepted in Germany</td></tr><tr><td>2, 3, 4, 9</td><td>4</td><td>In-company / school-based training</td></tr><tr><td>5, 11, 14, 15, 17, 19, 20</td><td>5</td><td>University of applied sciences</td></tr><tr><td>6, 12, 13, 16, 18, 21</td><td>6</td><td>University</td></tr><tr><td>10</td><td>6</td><td>Other exams</td></tr><tr><td>23</td><td>7</td><td>Undergraduate studies</td></tr><tr><td>24</td><td>7</td><td>Secondary/additional studies</td></tr><tr><td>25</td><td>7</td><td>Other (promotion-) advanced training for graduates</td></tr></table>	Ausbildung	ausbildung_agg	Value label	1, 22	1	Without (recognised) vocational training	7	2	Vocational training not accepted in Germany	8	3	University degree not accepted in Germany	2, 3, 4, 9	4	In-company / school-based training	5, 11, 14, 15, 17, 19, 20	5	University of applied sciences	6, 12, 13, 16, 18, 21	6	University	10	6	Other exams	23	7	Undergraduate studies	24	7	Secondary/additional studies	25	7	Other (promotion-) advanced training for graduates
Ausbildung	ausbildung_agg	Value label																																
1, 22	1	Without (recognised) vocational training																																
7	2	Vocational training not accepted in Germany																																
8	3	University degree not accepted in Germany																																
2, 3, 4, 9	4	In-company / school-based training																																
5, 11, 14, 15, 17, 19, 20	5	University of applied sciences																																
6, 12, 13, 16, 18, 21	6	University																																
10	6	Other exams																																
23	7	Undergraduate studies																																
24	7	Secondary/additional studies																																
25	7	Other (promotion-) advanced training for graduates																																
Notes on quality	<div>1) BeH</div> <div>“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 and Koch 2004, p. 63).</div> <div>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%.</div> <div>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 and Koch, 2004, p. 63).</div> <div>For the variable “Vocational training (imputed)” (ausbildung_imp, see Section 5.5.9), 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.</div> <div>2) ASU</div>																																	

	<p>As a result of the switchover from coArb to VerBIS it is not possible to distinguish correctly between “no completed vocational training” and “no information available” in the ASU data source between 2006 and 2008. A missing value in this period therefore does not necessarily mean that the person has no vocational training or that there is no information available on vocational education and training, but may also mean that it was not possible to apply the relevant data generation procedure.</p> <p>2) XASU</p> <p>The degree of completeness in the XASU data source is generally low.</p>
--	---

5.5.9 Vocational training (imputed) (ausbildung_imp)

Variable label	Vocational training (imputed)
Variable name	ausbildung_imp
Category	personal variables
Origin	BeH
Data type	numerical
Detailed description	<p>The variable Vocational training (imputed) is a supplement to the variable Vocational training (ausbildung) and contains additional and harmonized information on the vocational training of employees for BeH spells. The variable thus offers a solution to the problems concerning the variable Vocational training (ausbildung) described in Section 5.5.8. The imputation procedure is described in Thomsen et al (2018), which is based on the work of Fitzenberger et al. (2006).</p> <p>As the variable only uses the training information from BeH notifications and because the educational categories of the old and the new occupation codes had to be harmonised for the variable Vocational training (ausbildung), the variable ausbildung_imp has other categories than the variable ausbildung.</p>

5.5.10 School leaving qualification (schule)

Variable label	School leaving qualification
Variable name	schule
Category	personal variables
Origin	BeH, ASU, XASU
Data type	numerical
Detailed description	<p>This variable contains the school leaving qualification. Different values are possible depending on the source.</p> <p>1) BeH</p> <p>With the switch to the new occupation code (see Section 4.2) the possible values of the variable change. The values from the old occupation code are:</p> <p>5 Grade-/lower school certificate, intermediate school or equivalent qualification</p> <p>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</p> <p>9 Upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling</p> <p>With the new occupation code the values are:</p> <p>1 No school leaving certificate</p>

	<p>4 Lower secondary school certificate/ grade school certificate</p> <p>6 Intermediate school leaving certificate</p> <p>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</p> <p>2) ASU, XASU</p> <p>The following values are possible for these data sources:</p> <p>1 No school leaving certificate</p> <p>4 Lower secondary school certificate/ grade school certificate</p> <p>6 Intermediate school leaving certificate</p> <p>7 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college</p> <p>9 Upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling</p> <p>They are valid at the beginning of the period of job-search or participation in a measure. In the case of people seeking an apprenticeship position, the variable may also contain the school qualification they are working towards in the XASU data source.</p>
Notes on quality	<p>The degree of completeness in the BeH has been decreasing continuously over time and seems to have levelled off at under 2/3 in recent years. In the XASU it has been increasing continuously and has levelled off at over 2/3 since 2013 (XASU) 2012. In the ASU the degree of completeness is generally high.</p>

5.6 Information on employment, benefit receipt and job search

5.6.1 Daily wage, daily benefit rate (tentgelt)

Variable label	Daily wage/daily benefit
Variable name	tentgelt
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH
Data type	numerical
Detailed description	<p>1) BeH</p> <p>In BeH observations, this variable shows the employee's gross daily wage. It is calculated from the fixed-period wages reported by the employer and the duration of the (unsplit) original notification period in calendar days. The daily wage is shown in euros.</p> <p>Until 1998, employers in principle only reported the earnings which were subject to social security contributions. Earnings below the marginal part-time income threshold were not reported. 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.</p> <p>Since the inclusion of marginal part-time employees in the employment notification procedure on 1 April 1999, earnings below the marginal part-time income threshold have also been recorded; the upper earnings limit still applies as the upper ceiling. In some cases, however, the reported earnings nonetheless exceed the upper earnings limit. Generally, this can probably be attributed to the payment of annual bonuses which the employer can add to the regular earnings in the annual,</p>

	<p>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 information, however, may be considered less error-prone due to its insurance relevance.) The marginal part-time income threshold and the upper earnings limit for statutory pension insurance differ from year to year as well as between eastern and western Germany (the decisive factor is the location of the establishment). An overview of these limits and thresholds can be found under http://fdz.iab.de.</p> <p>A daily wage reported as 0 euros can be put down to “employment interruption notifications”. During these periods, the employment relationship continues to exist in legal terms, but without pay. This is the case for periods of illness after the end of continued payment of wages, for periods of maternity leave and for sabbaticals.</p> <p>From 2013 onwards, the number of notifications with a reason for deregistration of 54 (notification of a one-off wage) increases sharply (see Section 5.6.11). 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.</p> <p>The daily wage is shown with two decimal places. All values greater than 0 and smaller than 0.01 were rounded up to 0.01. This makes it possible to identify the above-mentioned employment interruption notifications with the condition daily wage = 0.</p> <p>2) LeH</p> <p>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 1 January 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 1 January 1998 onwards it applies to calendar days.</p> <p>Since 1 January 2005, a daily benefit rate reported as 0 euros can be put down to benefit suspension periods or interruptions of benefit payments. If a reason for end of benefit is reported for an observation with a daily benefit rate equal to 0, then it is a notification of interruption of benefit payments. In the case of observations that reflect a period of benefit suspension, the entitlement is the same as before the start of the benefit suspension period.</p>
--	--

5.6.2 Occupation - current/most recent (KldB 1988) (beruf)

Variable label	Occupation – current/most recent (KldB 1988)
Variable name	beruf
Category	information on employment, benefit receipt and job search
Origin	BeH, ASU, XASU
Data type	numerical
Detailed description	<p>1) BeH</p> <p>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 Bundesanstalt für Arbeit, 2005, p. V).</p>

	<p>For this, the employer encodes the employee's job in accordance with the "Classification of Occupations. Systematic and Alphabetical Directory of Job Titles" (published by the Federal Employment Agency, Nuremberg, 1988), which contains approx. 25,000 job titles. The occupational classification consists of a 3-digit code and comprises about 330 values.</p> <p>Employment notifications with an end date later than 30 November 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.</p> <p>2) ASU, XASU The variable contains the occupation of the last job. See 1) with regard to the occupation code.</p>
Notes on quality	<p>1) 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.</p> <p>2) ASU, MTH The degree of completeness decreases significantly from 2014 to 2017.</p> <p>3) XASU The occupation variable is not filled for almost the entire period available.</p>

5.6.3 Occupational group - current/most recent (KldB 2010) (beruf2010_3)

Variable label	Occupational group - current/most recent (KldB 2010), 3-digit
Variable name	beruf2010_3
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, ASU, XASU
Data type	numerical
Detailed description	<p>1) 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 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). 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 1300 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). Employment notifications with an end date earlier than 30 November 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.</p> <p>2) LeH, ASU, XASU The variable contains the occupation of the last job. See 1) with regard to the occupation code.</p>

Notes on quality	<p>1) 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.</p> <p>2) LeH This variable was previously not filled in the LeH, since the occupational data in that source was of poor quality. By now, the occupation in the LeH is adopted from the job seeker history, leading to an increase in quality.</p> <p>3) XASU The variable is not filled for almost the entire period available.</p>
------------------	--

5.6.4 Occupational sub-group - current/most recent (KldB 2010) (beruf2010_4)

Variable label	Occupational sub-group - current/most recent (KldB 2010), 4-digit
Variable name	beruf2010_4
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, ASU, XASU
Data type	numerical
Detailed description	<p>1) 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 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). 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 1300 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). Employment notifications with an end date earlier than 30 November 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.</p> <p>2) LeH, ASU, XASU The variable contains the occupation of the last job. See 1) with regard to the occupation code.</p>
Notes on quality	<p>1) 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 the reunification.</p> <p>2) LeH This variable was previously not filled in the LeH, since the occupational data in that source was of poor quality. By now, the occupation in the LeH is adopted from the job seeker history, leading to an increase in quality.</p> <p>3) XASU</p>

	The variable is not filled for almost the entire period available.
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.

5.6.5 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, benefit receipt and job search
Origin	BeH, LeH, ASU, XASU
Data type	numerical
Detailbeschreibung	<p>1) 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 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).</p> <p>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 1300 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).</p> <p>Employment notifications with an end date earlier than 30 November 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.</p> <p>2) LeH, ASU, XASU The variable contains the occupation of the last job. See 1) with regard to the occupation code.</p>
Notes on quality	<p>1) 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 the reunification.</p> <p>2) LeH This variable was previously not filled in the LeH, since the occupational data in that source was of poor quality. By now, the occupation in the LeH is adopted from the job seeker history, leading to an increase in quality.</p> <p>3) XASU The variable is not filled for almost the entire period available.</p>

5.6.6 Part-time (teilzeit)

Variable label	Part-time
Variable name	teilzeit
Category	information on employment, benefit receipt and job search

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 and Thomsen (2016). A similar accumulation of missing values occurs in 1991 due to the reunification. No imputation is performed here.

5.6.7 Employment status (erwstat)

Variable label	Employment status
Variable name	erwstat
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, LHG, ASU, XASU
Data type	numerical
Detailed description	<p>This variable takes on different values with different meanings for each data source.</p> <p>1) BeH For BeH observations, the variable ‘employment status’ corresponds to the person group recorded in the new notification procedure (DEÜV) from 1 January 1999 onwards. It indicates contribution- or benefit-related particularities of the employment relationship. 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. 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. 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 vocational education and training', 'occupational status and working hours' and 'occupation' variables as well as other information. In many cases, however, conclusive allocations are not possible. Since 1 April 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.</p> <p>2) LeH</p>

	<p>For LeH observations, the variable 'employment status' 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.</p> <p>3) LHG</p> <p>For LHG spells, the variable shows whether the person is underage and able to work, adult and able to work, or unable to work and beyond the retirement pension limit.</p> <p>Since the reporting logic would make it possible to re-identify the exact date of birth in many cases, the original dates were changed by means of the anonymization procedure described in Section 3.2.6.</p> <p>4) ASU, XASU</p> <p>For ASU/XASU observations, the 'employment status' variable reports the job search status.</p> <p>A distinction is made between those who are "unemployed and seeking work", "not unemployed and seeking work", "seeking advice" and "not seeking work". Applicants who only want advice from the BA are considered as "seeking advice". In addition, there are rehabilitants and, before 2008, persons aged 58 or older who are not fully available for placement. As of 1 August 2016, this may also include persons who no longer require support, but who are nevertheless still under the job center's care.</p> <p>"Not seeking work" mainly subsumes persons of whom activation or placement cannot be reasonably expected according to § 10 SGB II. Similarly, persons with an incapacity to work of more than 42 days who continue to receive ALG II are listed in the system under this status.</p> <p>Since the reporting logic would make it possible to re-identify the exact date of birth in many cases, the original dates were changed by the anonymization procedure described in Section 3.2.6.</p>
Notes on quality	<p>1) LHG</p> <p>In the LHG it can be observed that there is an above-average number of 15-year-olds and to a lesser extent 16- and 17-year-olds classified as unable to work. 15- and 16-year-old benefit recipients of the authorised municipalities may therefore be under-represented, as 'individuals who are unable to work' are not included in the IEB.</p> <p>In some cases, there may be conflicting information on a person's ability to work if he or she is a member of different benefit units (BGs) at the same time. A possible reason for this is the determination of the ability to work by the job center specialist supervising the BG or by the respective responsible medical services. They can come to different results or document these at different times. A further reason may be different birthday records for simultaneous BG customers, which may affect the status of their ability to work.</p> <p>2) ASU, XASU</p> <p>The categories "seeking advice" and "not seeking work" have only existed since the introduction of Verbis (see Section 4.5). Due to late notifications, however, entries before 2006 can also be found. The characteristic "not seeking work" is considered to be under-recorded before 2008. In the XASU, the quality of status determination is partly limited for some institutions, especially in the first years.</p>

5.6.8 Transition zone (gleitz)

Variable label	Transition zone
Variable name	gleitz
Category	information on employment, benefit receipt and job search
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 1 April 2003.

5.6.9 Temporary agency work (leih)

Variable label	Temporary agency work
Variable name	leih
Category	information on employment, benefit receipt and job search
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 30 November 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%.

5.6.10 Fixed-term contract (befrist)

Variable label	Fixed-term contract
Variable name	befrist
Category	information on employment, benefit receipt and job search
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 30.11.2011.
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%.

5.6.11 Reason of cancellation/notification/termination (grund)

Variable label	Reason of cancellation/notification/termination
Variable name	grund
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, LHG, ASU, XASU
Data type	numerical
Detailed description	1) BeH

	<p>In BeH observations, the 'reason for 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.</p> <p>The reasons for submitting employment notifications are encoded according to the regulations of the notification procedure, which has been in effect since 1 January 1999 (in accordance with DEÜV).</p> <p>2) LeH</p> <p>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 about the reasons for the start of the benefit receipt, since the LeH is filled with the notifications from the employment agencies to the health insurance about completed benefit receipt durations.</p> <p>3) LHG</p> <p>The LHG observations contain the ‘reason for discontinuation of Unemployment Benefit II’ and indicate the reason why current benefits have been discontinued. The ‘reason for discontinuation of Unemployment Benefit II’ variable refers to the individual, not to the benefit unit. If the Unemployment Benefit II receipt of a different member of the benefit unit is discontinued, new observations for all members of the benefit unit are started on this date, but the reason for discontinuation of Unemployment Benefit II is only available for the individual whose benefit is discontinued. This variable is valid exactly at the end of the original observation.</p> <p>4) ASU</p> <p>In the case of ASU observations, the variable contains the deregistration or exit reason. In the case of a change of legal sphere, the observation is split artificially and ‘generated by data splitting’ is entered as the reason for deregistration. In order to depict the reasons for deregistration correctly it is also necessary to take into account the variable ‘status after job search’.</p> <p>The number of values of the variable was reduced from 26 April 2003 onwards. For analyses over long periods of time, the old values can be recoded to the currently valid ones using the table below (in the 5000s number range):</p> <table><tr><td>29->60</td><td>33->60</td><td>37->66</td><td>42->65</td><td>46->67</td><td>50->75</td><td>54->78</td></tr><tr><td>30->60</td><td>34->60</td><td>38->66</td><td>43->70</td><td>47->67</td><td>51->74</td><td></td></tr><tr><td>31->61</td><td>35->60</td><td>39->71</td><td>44->74</td><td>48->78</td><td>52->76</td><td></td></tr><tr><td>32->60</td><td>36->61</td><td>40->69</td><td>45->77</td><td>49->69</td><td>53->68</td><td></td></tr></table> <p>5) XASU</p> <p>In the case of XASU observations, the variable contains the deregistration or exit reason. In the case of a change of legal sphere, the observation is split artificially and ‘generated by data splitting’ is entered as the reason for deregistration.</p>	29->60	33->60	37->66	42->65	46->67	50->75	54->78	30->60	34->60	38->66	43->70	47->67	51->74		31->61	35->60	39->71	44->74	48->78	52->76		32->60	36->61	40->69	45->77	49->69	53->68	
29->60	33->60	37->66	42->65	46->67	50->75	54->78																							
30->60	34->60	38->66	43->70	47->67	51->74																								
31->61	35->60	39->71	44->74	48->78	52->76																								
32->60	36->61	40->69	45->77	49->69	53->68																								
Notes on quality	<p>1) BeH</p>																												

	<p>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.</p> <p>2) LHG</p> <p>The degree of completeness for the reason for notification in the LHG data sources is very small (< 20%) across all years.</p>
--	--

5.6.12 Employment status prior to job search (estatvor)

Variable label	Employment status prior to job search
Variable name	estatvor
Category	information on employment, benefit receipt and job search
Origin	ASU, XASU
Data type	numerical
Detailed description	<p>For ASU and XASU observations, this variable shows the employment status prior to the job search activity. From December 2002 onwards the number of values of the variable was reduced substantially. The values of older observations were re-coded to the currently valid values, which are thus valid for the entire period. If an observation within the ASU/XASU is split artificially (for instance in the case of a change of legal sphere from SGB III to SGB II), the reason for registration is reported as 'generated by data splitting'. However, this does not apply for episode splitting within the LPP-ADIAB.</p> <p>This information does not refer to the start date of the episode, but to the start of the original time period.</p>
Note on quality	The proportion of valid values (degree of completeness) increases for XASU observations from approx. 8 % to approx. 78 % between 2005 and 2013.

5.6.13 Employment status after job search (estatnach)

Variable label	Employment status after job search
Variable name	estatnach
Category	information on employment, benefit receipt and job search
Origin	ASU
Data type	numerical
Detailed description	<p>The variable contains the person's status after leaving unemployment. Longer periods of illness can be identified via this variable.</p> <p>The values are classified as follows:</p> <ul style="list-style-type: none"> • 1000s: measure (assisted employment) • 2000s: non-assisted employment • 3000s: training etc. • 4000s: self-employment • 5000s: exclusion • 6000s: other

5.6.14 Client profile (profil)

Variable label	Client profile
----------------	----------------

Variable name	profil
Category	information on employment, benefit receipt and job search
Origin	ASU
Data type	numerical
Detailed description	<p>The variable reports the client profile assigned to the client in the profiling process. The profiling process serves to create a client profile, i.e. a list of the client's skills, experiences and interests with labour-market relevance, in order to identify the client's position in the labour market more easily. Towards the end of the profiling process, the items are summarised to create a client profile. To this end, the client's overall integration prospects are first ascertained. The following options are available:</p> <ul style="list-style-type: none"> • good integration prospects (integration into the regular labour market within 12 months is realistic) • complex (integration into the regular labour market within 12 months is not realistic) • other <p>The allocation of the client profile depends on the identification of the integration prospects. Clients whose integration prospects are classed as good can be assigned the client profiles 'market profile', 'activation profile' and 'assistance profile', while clients with complex prospects are assigned the client profiles 'development profile', 'stabilisation profile' or 'support profile'. The selection of the specific client profile is based on the need for action as assessed by the placement officer. If the client's prospects are classed as 'other', the option 'assignment not required' or – only for SGB II clients – 'integrated but in receipt of benefits' may be selected as the client profile.</p>
Notes on quality	The variable was introduced in 2009 but was mapped back to 2006 using other variables. The quality has been assessed as reliable by the BA statistics department since 2010.

5.6.15 Type of termination of last job (art_kuend)

Variable label	Reason for end of previous employment
Variable name	art_kuend
Category	information on employment, benefit receipt and job search
Origin	ASU
Data type	numerical
Detailed description	This variable describes how the last employment or training relationship was terminated before a period of job search. It can therefore be used to identify job-to-job placements.

5.6.16 Desired working hours of the job sought (arbeitszeit)

Variable label	Working hours of job application
Variable name	arbeitszeit
Category	information on employment, benefit receipt and job search
Origin	ASU
Data type	numerical
Detailed description	During the placement procedure, jobseekers indicate how many working hours the job they are seeking should have.

5.6.17 Residual claim/planned duration (restanspruch)

Variable label	Residual claim/planned duration
----------------	---------------------------------

Variable name	restanspruch
Category	information on employment, benefit receipt and job search
Origin	LeH
Data type	numerical
Detailed description	<p>The variable has a different meaning depending on the data source.</p> <p>1) LeH</p> <p>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. If the end date of the benefit receipt is before 1 January 1998, the remaining entitlement is reported in working days, after this date it is reported in calendar days. This information does not refer to the start date of the episode, but to the start of the original time period.</p>
Notes on quality	<p>1) LeH</p> <p>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.</p> <p>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.</p>

5.6.18 Type of provider (traeger)

Variable label	Type of provider
Variable name	traeger
Category	information on employment, benefit receipt and job search
Origin	LHG, ASU, XASU
Data type	numerical
Detailed description	<p>The variable contains the type of institution responsible for implementing Unemployment Benefit II (LHG), the type of institution providing the measure (MTH/XMTH) or the institution responsible for managing the applicant profile (ASU/XASU). The variable contains not only the "Bundesagentur für Arbeit (BA)" as the type of institution responsible for implementing Social Code Book III (SGB III), but also three types of institution responsible for implementing Social Code Book II (SGB II). For further information see Section 2.4.3. The type of institutions responsible for implementing SGB II may change over time (e.g., from a cooperation</p>

	<p>of an employment agency and a municipality to a municipality exercising its duties separately, or from joint facilities to an authorised municipality). It must also be taken into account that the district territories covered by the institutions are not always distinct and thus may not necessarily correspond to the boundaries of the districts. The same applies for the employment agencies.</p> <p>ASU and XASU observations contain the type of institution that holds the records of the applicant pool data.</p>
--	---

5.6.19 Start date of unemployment (alo_beg)

Variable label	Start date of unemployment
Variable name	alo_beg
Category	information on employment, benefit receipt and job search
Origin	LeH, LHG, ASU, XASU
Data type	numerical
Detailed description	<p>The variable reports the start date of an uninterrupted sequence of periods of unemployment and is valid at the beginning of the observation.</p> <p>The following gaps do not result in an interruption of the period of unemployment:</p> <ul style="list-style-type: none"> any gap lasting seven days or less periods of illness lasting up to 42 days

5.6.20 Duration of unemployment (alo_dau)

Variable label	Duration of unemployment
Variable name	alo_dau
Category	information on employment, benefit receipt and job search
Origin	LeH, LHG, ASU, XASU
Data type	numerical
Detailed description	<p>The variable reports the duration (in days) of an uninterrupted sequence of periods of unemployment and is valid at the beginning of the observation.</p> <p>The following gaps do not result in an interruption of the period of unemployment:</p> <ul style="list-style-type: none"> any gap lasting seven days or less periods of illness lasting up to 42 days <p>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".</p> <p>Prior to 1997, the value "0" does not mean that the individual was not unemployed, as the ASU/XASU sources are not available here.</p>

5.7 Location data

5.7.1 Place of residence - district (Kreis) (wo_kreis)

Variable label	Place of residence - district (Kreis)
Variable name	wo_kreis
Category	location data
Origin	BeH, LeH, LHG, ASU, XASU
Data type	numerical

Hierarchy	federal state district
Detailed description	<p>In BeH and LeH observations, the place of residence at 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.</p> <p>In the BeH, the place of residence is determined at the end of each year and added consistently to all datasets of a year. For the LHG, XLHG and XASU sources, the place of residence applies to the whole period of the original observation. For the ASU and LeH, the variable contains the place of residence at the beginning of the original period of time. 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.</p> <p>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 31 December 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 31 December 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.</p>
Notes on quality	<p>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.</p> <p>In the year 2015, the data show a reduction in the change of residence of approx. 10-15% 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.</p>
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 (wo_bula) is provided.

5.7.2 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, LHG, ASU, XASU
Data type	numerical
Hierarchy	federal state district
Detailed description	<p>This variable is an aggregation of the "district" variable to the 16 German federal states.</p> <p>In 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 under Place of residence: district (Kreis) (wo_kreis).</p>

5.7.3 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, LHG, XLHG, ASU, XASU
Data type	numerical
Hierarchy	regional directorate employment agency
Detailed description	<p>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 and ASU 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.</p> <p>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 31 December 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 31 December 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. 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 1 July 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).</p>
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.

5.7.4 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, LHG, ASU, XASU
Data type	numerical
Hierarchy	regional directorate employment agency
Detailed description	This variable is an aggregation of the variable 'place of residence: employment agency' at the level of the regional directorates. Further information can be found there.

5.8 Establishment variables

5.8.1 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	<p>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. 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”) (Bundesagentur für Arbeit, 2009).. 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.</p> <p>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.</p>

5.8.2 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	<p>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”) (Statistisches Bundesamt, 2003). The WZ93 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 1 („Nomenclature générale 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“).</p> <p>Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.</p>
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).

5.8.3 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	<p>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”) (Statistisches Bundesamt, 2002). The WZ93 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 1 („Nomenclature générale 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“).</p> <p>Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.</p>

5.8.4 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	<p>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.) (Statistisches Bundesamt, 2003). 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.</p> <p>Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.</p>
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).

5.8.5 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.) (Statistisches Bundesamt, 2003). 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. Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.
----------------------	--

5.8.6 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.) (Statistisches Bundesamt, 2008). The WZ08 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 2. 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).

5.8.7 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.) (Statistisches Bundesamt, 2008). The WZ08 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 2. Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.

5.8.8 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	<p>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).</p> <p>Further information on the WS73 classification can be found in the description of variable w73_3.</p>

5.8.9 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	<p>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.</p> <p>A detailed description of the procedure can be found in Eberle et al. (2011).</p>

5.8.10 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	<p>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).</p> <p>Further information on the WS93 classification can be found in the description of variable w93_3.</p>

5.8.11 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. A detailed description of the procedure can be found in Eberle et al. (2011).

5.8.12 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). Further information on the WZ08 classification can be found in the description of variable w08_3.

5.8.13 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. A detailed description of the procedure can be found in Eberle et al. (2011).

5.8.14 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

	<p>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.</p> <p>An establishment does not necessarily have to be included in the BHP in the year of its first appearance, since 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.</p>
--	---

5.8.15 First appearance (grd_dat)

Variable label	first appearance
Variable name	grd_dat
Category	establishment variables
Origin	BHP
Data type	date
Detailed description	<p>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.</p> <p>An establishment does not necessarily have to be included in the BHP in the year of its first appearance, since 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.</p>
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).

5.8.16 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	<p>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 2008, 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 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).</p> <p>An establishment does not necessarily have to be included in the BHP in the year of its last appearance, since 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.</p>

5.8.17 Last appearance (lzt_dat)

Variable label	last appearance
Variable name	lzt_dat

Category	establishment variables
Origin	BHP
Data type	date
Detailed description	<p>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 2008, 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).</p> <p>An establishment does not necessarily have to be included in the BHP in the year of its last appearance, since 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.</p>
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).

5.8.18 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 30 June 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.

5.8.19 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 30 June 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.

5.8.20 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.
----------------------	---

5.8.21 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	<p>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.</p> <p>The values are reported in euros for all years.</p> <p>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 cut.</p> <p>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.</p>

5.8.22 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	<p>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.</p> <p>In order to guarantee consistent regional allocations across the entire observation period, the information on the district was recoded to the territorial allocation of 31 December 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 31 December 2017. As the district boundaries have changed over time, cases would occur in which the</p>

	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.

5.8.23 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

- Antoni, Manfred; Schmucker, Alexandra; Seth, Stefan; vom Berge, Philipp (2019): Sample of Integrated Labour Market Biographies (SIAB) 1975–2017. FDZ Datenreport, 02/2019 (de), Nürnberg, 74 S.
- Bellmann, Lutz; Bender, Stefan; Bossler, Mario; Stephani, Jens; Wolter, Stefanie; Sliwka, Dirk; Kampkötter, Patrick; Laske, Katharina; Steffes, Susanne; Mohrenweiser, Jens; Nolte, André (2014): Arbeitsqualität und wirtschaftlicher Erfolg. Längsschnittstudie in deutschen Betrieben. Erster Zwischenbericht im Projekt. (Bundesministerium für Arbeit und Soziales. Forschungsbericht Arbeitsmarkt, 442), Berlin, 161 S.
- Bender, Stefan; Hilzendegen, Jürgen; Rohwer, Götz; Rudolph, Helmut (1996): Die IAB-Beschäftigtenstichprobe 1975-1990. Beiträge zur Arbeitsmarkt- und Berufsforschung 197, Nürnberg.
- 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: <http://statistik.arbeitsagentur.de/Statistischer-Content/Grundlagen/Methodenberichte/Beschaeftigungsstatistik/Generische-Publikationen/Methodenbericht-Neue-Erhebungsinhalte-Arbeitszeit-ausgeuebte-Taetigkeit-sowie-Schul-und-Berufsabschluss-in-der-Beschaeftigungsstatistik.pdf> (Zugriff: 31.03.2016).
- Broszeit, Sandra; Wolter, Stefanie (2015): LPP – Linked Personnel Panel * Arbeitsqualität und wirtschaftlicher Erfolg: Längsschnittstudie in deutschen Betrieben (Datendokumentation der ersten Welle). FDZ-Datenreport, 01/2015 (de), Nürnberg, 137 S.
- Broszeit, Sandra; Grunau, Philipp; Wolter, Stefanie (2016): LPP - Linked Personnel Panel 1415 * Arbeitsqualität und wirtschaftlicher Erfolg: Längsschnittstudie in deutschen Betrieben (Datendokumentation der zweiten Welle). FDZ-Datenreport, 06/2016 (de), Nürnberg, 202 S.
- Bundesagentur für Arbeit (Hrsg.) (2005): Schlüsselverzeichnis für die Angaben zur Tätigkeit in den Meldungen zur Sozialversicherung. Ausgabe Januar 2005, Nürnberg.
- 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: 12 June 2019)
- Bundesagentur für Arbeit (2011): Klassifikation der Berufe 2010. Band 1: Systematischer und alphabetischer Teil mit Erläuterungen, Nürnberg. URL: <http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB2010/Printausgabe-KldB-2010-Nav.html> (Zugriff: 21.04.2016).

- Bundesanstalt für Arbeit (1988): Klassifikation der Berufe – Systematisches und Alphabetisches Verzeichnis der Berufsbenennung, Nürnberg. URL: <http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB1975-1992/KldB1975-1992-Nav.html> (Zugriff: 21.04.2016)
- Cramer, Ulrich (1985): Probleme der Genauigkeit der Beschäftigtenstatistik. In: Allgemeines Statistisches Archiv 69: S. 56-68.
- Eberle, Johanna; Jacobebbinghaus, Peter; Ludsteck, Johannes; Witter, Julia (2011): Generation of time-consistent industry codes in the face of classification changes * Simple heuristic based on the Establishment History Panel (BHP). FDZ-Methodenreport, 05/2011 (en), Nürnberg, 21 S.
- Fitzenberger, Bernd; Osikominu, Aderonke; Völter, Robert (2006): Imputation rules to improve the education variable in the IAB employment subsample. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 126, H. 3, S. 405-436.
- Ludsteck, Johannes; Thomsen, Ulrich (2016): Imputation of the Working Time Information for the Employment Register Data. FDZ Methodenreport 01/2016 (en), Nürnberg.
- Mackeben, Jan; Ruf, Kevin; Grunau, Philipp & Wolter, Stefanie (2018): LPP - Linked Personnel Panel 1617 * Arbeitsqualität und wirtschaftlicher Erfolg: Längsschnittstudie in deutschen Betrieben (Datendokumentation der zweiten Welle). FDZ-Datenreport, 04/2018 (de), Nürnberg, 216 S.
- Meinken, Holger; Koch, Iris (2004): BA-Beschäftigtenpanel 1998-2002. Codebuch, Nürnberg.
- Paulus, Wiebke; Matthes, Britta (2013): The German Classification of Occupations 2010—Structure, Coding and Conversion Table. FDZ-Methodenreport, 08/2013 (en), Nürnberg.
- Schmucker, Alexandra; Eberle, Johanna; Ganzer, Andreas; Stegmaier, Jens; Umkehrer, Matthias (2018): Establishment History Panel 1975-2016. FDZ-Datenreport, 01/2018 (en), Nürnberg.
- Schütz, Holger; Gilberg, Reiner; Dickmann, Christian; Schröder, Helmut (2015): IAB-Beschäftigtenbefragung * Projekt "Arbeitsqualität und wirtschaftlicher Erfolg: Panelstudie zu Entwicklungsverläufen in deutschen Betrieben – Personenbefragung". FDZ-Methodenreport, 02/2015 (de), Nürnberg, 40 S.
- Schütz, Holger; Gilberg, Reiner; Knerr, Petra; Kellerhoff, Jette; Dickmann, Christian (2016): IAB-Beschäftigtenbefragung * Projekt "Arbeitsqualität und wirtschaftlicher Erfolg: Panelstudie zu Entwicklungsverläufen in deutschen Betrieben" – Personenbefragung. 2. Erhebungswelle. FDZ-Methodenreport, 02/2016 (de), Nürnberg, 43 S.
- Schütz, Holger; Knerr, Petra; Metselaar, Carolin; Middendorf, Lena (2017): IAB-Beschäftigtenbefragung * Projekt "Arbeitsqualität und wirtschaftlicher Erfolg: Panelstudie zu Entwicklungsverläufen in deutschen Betrieben" – Personenbefragung. 3. Erhebungswelle. (FDZ-Methodenreport, 11/2018 (de)), Nürnberg, 65 S.
- 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> (Zugriff: 12.06.2019).
- Statistisches Bundesamt (2003): Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003), Wiesbaden. URL: <https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftsklassifikationen/Content75/KlassifikationWZ2003.html> (Zugriff: 12.06.2019).

- 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> (Zugriff: 12.06.2019).
- Statistisches Bundesamt (2019): Staats- und Gebietssystematik, Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschlüssel_pdf.pdf?__blob=publicationFile&v=5 (Zugriff: 12.06.2019).
- 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.
- Tschersich, Nikolai; Gensicke, Miriam (2015): Vertiefende Betriebsbefragung "Arbeitsqualität und wirtschaftlicher Erfolg" 2012. FDZ-Methodenreport, 03/2015 (de), Nürnberg, 52 S.
- Tschersich, Nikolai; Gensicke, Miriam (2016): Vertiefende Betriebsbefragung "Arbeitsqualität und wirtschaftlicher Erfolg" 2014 * Welle 2. FDZ-Methodenreport, 03/2016 (de), Nürnberg, 59 S.
- Tschersich, Nikolai; Gensicke, Miriam (2016): Vertiefende Betriebsbefragung „Arbeitsqualität und wirtschaftlicher Erfolg“ 2016 –Welle 3. FDZ-Methodenreport 08/2018 (de), Nürnberg.
- Wermter, Winfried; Cramer, Ulrich (1988): Wie hoch war der Beschäftigtenanstieg seit 1983? – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 4/88, S. 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 <http://fdz.iab.de/en.aspx>.

7.2 List of abbreviations

AA	Agentur für Arbeit / Arbeitsamt	employment agency / employment office
ALG	Arbeitslosengeld	unemployment benefit
ARGE	Arbeitsgemeinschaft	cooperation of employment agencies and municipalities
ASU	Arbeitsuchendenhistorik	Jobseeker History
A2LL	Arbeitslosengeld II – Leistungen zum Lebensunterhalt	unemployment benefit II - benefits to secure a livelihood
BA	Bundesagentur für Arbeit	Federal Employment Agency
BeH	Beschäftigtenhistorik	Employee History
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
MTH	Maßnahmeteilnahmehistorik	Participants-in-Measures History File
NACE	Nomenclature générale des activités économiques dans les communautés européennes	Nomenclature générale des activités économiques dans les communautés européennes
NUTS	Nomenclature des unités territoriales statistiques	Nomenclature des unités territoriales statistiques
SGB	Sozialgesetzbuch	German Social Code
SIAB	Stichprobe der Integrierten Arbeitsmarktbiographien	Sample of Integrated Labour Market Biographies
VerBIS	Vermittlungs- und Beratungsinformationssysteme	Information System for Placement and Counselling
XASU	Arbeitsuchendenhistorik aus XSozial-BA-SGB II	Jobseeker History from XSozial-BA-SGB II
XMTH	Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II	Participants-in-Measures History File from XSozial-BA-SGB II
ZEW	Zentrum für europäische Wirtschaftsforschung	Centre for European Economic Research
zKT	Zugelassene kommunale Träger	Authorised municipalities

8 Imprint

FDZ-Datenreport 07|2019 EN

Date of publication

22 October 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

All rights reserved

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

Download

http://doku.iab.de/fdz/reporte/2019/DR_07-19_EN.pdf

Documentation version

LPP-ADIAB_7517_EN_v1_dok1, DOI: 10.5164/IAB.FDZD.1907.en.v1

Dataset version

LPP-ADIAB 7517 v1, DOI: 10.5164/IAB.LPP-ADIAB7517.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

Jan Mackeben
Phone: +49 911 179-9294
Email: jan.mackeben@iab.de