FDZ-DATENREPORT

Documentation of labour market data

Dr. Sekou Keita, Dr. Parvati Trübswetter

IAB-BAMF-SOEP Survey of Refugees linked to administrative data of the IAB

Dr. Sekou Keita, Dr. Parvati Trübswetter (Institut für Arbeitsmarkt- und Berufsforschung)

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

FDZ 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

List of Figures

List of Tables

Zusammenfassung

Dieser Datenreport beschreibt die verknüpften Daten der IAB-BAMF-SOEP Befragung von Geflüchteten mit administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB).

Abstract

This data report describes the linked data of the IAB-BAMF-SOEP survey of refugees with administrative data of the Institute for Employment Research (IAB).

Keywords

Acknowledgement

Data availability

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

1 Introduction and outline

1.1 Introduction

This data report describes the linked survey data of the IAB-BAMF-SOEP sample of refugees with the administrative data of the Institute for Employment Research (IAB). These linked process and survey data consist of information collected in interviews from participants in the IAB-BAMF-SOEP survey of refugees and their personal data from the BA's administrative data, provided that they gave their consent to the linking of their data and that they could be identified in the administrative dataset. In order to analyze IAB-BAMF-SOEP-ADIAB, this FDZ data report alone is not sufficient, but should be seen as a complement to the existing documentation of IAB-BAMF-SOEP survey. Please additionally use the FDZ data reports on waves 1 to 3 of the IAB-BAMF-SOEP (see e.g. Brücker, Rother, and Schupp 2017).

The IAB-BAMF-SOEP survey of refugees is an annual repeated survey of people seeking protection in Germany and their household members, which was conducted for the first time in 2016. The sample of respondents was drawn from the Central Register of Foreigners (AZR) and is representative of asylum seekers who entered Germany between January 1, 2013 and December 31, 2016 and filed an asylum application by the end of June 2016 at the latest. The survey was extended in 2017 to add protection seekers who arrived by December 31, 2016 and were registered by January 1, 2017. Asylum seekers with a higher probability of remaining at the time of sampling were over-sampled due to the longitudinal design. In addition, women and persons over 30 years of age were also over-sampled. With appropriate weighting, it is nevertheless possible to make representative statements about refugees who entered Germany in the above-mentioned period.

In carrying out the project, the IAB cooperates with the Research Center of the Federal Office for Migration and Refugees (BAMF-FZ) and with the Socio-Economic Panel (SOEP) at the German Institute for Economic Research (DIW).

The interviews were conducted by trained interviewers from the KANTAR Public institute (formerly TNS Infratest Sozialforschung). The first wave of the survey, which took place in 2016, included 4,465 adults from 3,289 households. In the following year, 5,595 interviews were realized with adult persons from 3,822 households. Of these, 2,630 individuals participated in the survey for the second time. In 2018, 4,376 adults (including 510 new respondents) were interviewed in 3,061 households.

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)

These data, which come from different sources, are merged in the IEB and the statuses are updated daily.

While the Employee History (Beschäftigtenhistorik - BeH) is the origin of the information on employment subject to social security and marginal part-time employment, the receipt of benefits in accordance with Social Code Book III (SGB III) and Social Code Book II (SGB II) is recorded in the Benefit Recipient History (Leistungsempfängerhistorik - LeH) and the Unemployment Benefit II Recipient History (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 the SGB II, while participation in employment and training measures is recorded in the Participants-in-Measures History Files (Maßnahmeteilnahmehistoriken – MTH and XMTH). These participations were reported to the BA by approved municipal institutions via the transmission standard XSozial-BA-SGB II.

This Data report is structured as follows. In addition to the introduction, Section 1 contains information on data access as well as an outline of the data, the volume structure and a list of variables. A description of the individual data sources can be found in Section 2. Data preparation and data quality are discussed in Sections 3 and 4, while the individual variables are described in Section 5.

1.2 Data use

1.2.1 Data access

The IAB-BAMF-SOEP-ADIAB can only be accessed within the context of guest stays at the FDZ and subsequent remote data processing. 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 website.

1.2.2 Data management

The IAB-BAMF-SOEP-ADIAB dataset, which contains both English and German labels, is structured according to the data source of the variables and stored in several files. The IAB-BAMF-SOEP personal data serve as the starting point. They contain the personal characteristics of the IAB-BAMF-SOEP respondents collected in the interview as well as their personal and household identifiers. The identifiers can be used to add other partial data sets of the IAB-BAMF-SOEP survey data - such as household data or information on employment or unemployment benefit biographies.

On the one hand, the administrative data in IAB-BAMF-SOEP-ADIAB include the administrative personal file. This contains identifiers (personal ID and system-free firm identifiers), personal characteristics, information on employment, benefit receipt, job search and participation in labour market policy measures, characteristics on place of residence and technical characteristics. On the other hand, the administrative part of IAB-BAMF-SOEP-ADIAB consists of the establishment file. This contains a firm identifier, which allows a link with the administrative personal file, as well as year-related establishment characteristics. The variables marked with an "*" in the list of characteristics (see Table 4) are contained in the firm file. Thus, survey data and administrative firm or

personal characteristics are organizationally separated from each other, which makes the structure of the data clear and saves storage space in data management.

Another element of IAB-BAMF-SOEP-ADIAB is the linkage file, which contains the person number as well as a characteristic that indicates how the linkage with the administrative data was achieved. The file contains only the IAB-BAMF-SOEP respondents for whom administrative data can be provided. By linking the data with the survey data, attrition analyses can be carried out on the basis of all IAB-BAMF-SOEP participants.

Certain variables, which make it possible to identify individuals or establishments, are accessible 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:

Administrative 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)

It should be noted that the sensitive variables are usually already included in the data in an aggregated version. The sensitive variables are only disseminated in detail if the information in the aggregate state is insufficient to achieve the research objective.

Additional establishment variables can be provided upon request as blocks of variables from the Establishment History Panel (Betriebs-Historik-Panel - BHP) (see Figure 1). In addition, the extension files on Worker flows (inflows/outflows) and Establishment dynamics (entries/exits) can be provided upon request. Further information on the BHP is available at http://fdz.iab.de/en.aspx.

Figure 1: Data management of the IAB-BAMF-SOEP-ADIAB
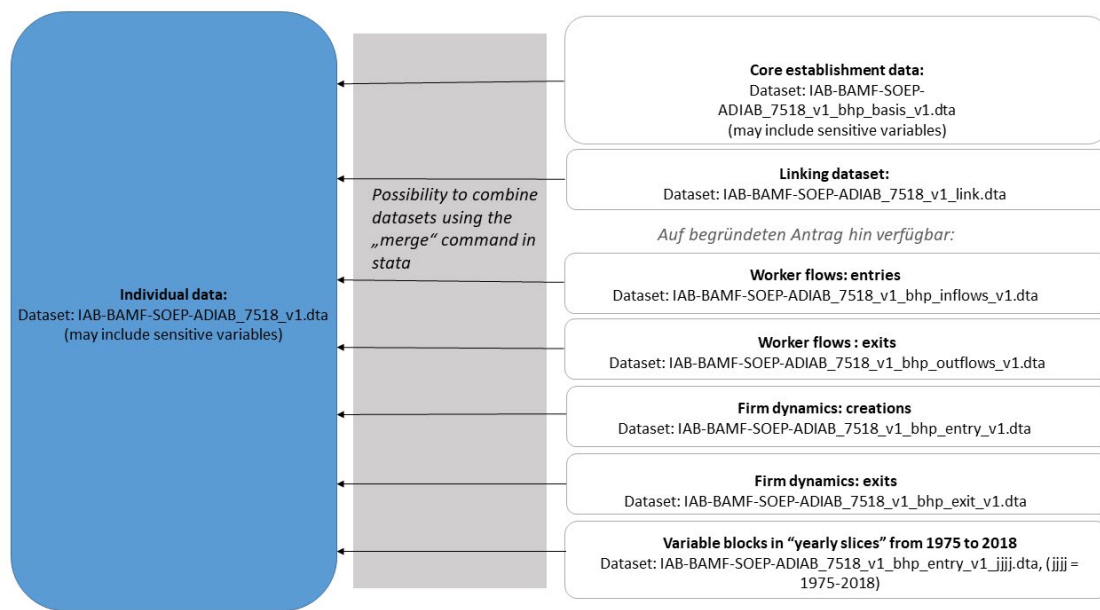
Table 1: File names of IAB-BAMF-SOEP-ADIAB data

Tracking	Original	Survey data	Generated data	Spell data
ppath.dta				
ppathl.dta				
hpath.dta	pl.dta	design.dta	pgen.dta	artkalen.dta
hpathl.dta	hl.dta		hgen.dta	biocouplm.dta
pbrutto.dta	biol.dta		bioage17.dta	biocouply.dta
hbrutto.dta	jugendl.dta		bioagel.dta	biomarsm.dta
hbrutt.dta			kidlong.dta	biomarsy.dta
pbr_exit.dta			pequiv.dta	lifespell.dta
			biobirth.dta	pbiospe.dta
			bioedu.dta	refugspell.dta
			bioimmig.dta	
			biojob.dta	
			bioparen.dta	
			biopupil.dta	
			bioresid.dta	
			biosib.dta	
			biosoc.dta	
			biotwin.dta	
			cognit.dta	
			cog_refu.dta	
			health.dta	
			mihinc.dta	
			pflege.dta	
			pkal.dta	

1.3   Profile

Table 2:   Summary description

Categories	Notes
Topics/ Groups of variables	<u>Survey data:</u> Origin and route to Germany Status of asylum procedure Accommodation in Germany Material resources and social participation Awareness, need and use of support and counseling services Language skills and language course attendance School, university and vocational training abroad and in Germany Employment and income abroad and in Germany Health status Personality Attitudes and values Social networks Family situation Children's participation in education Consent form for register linkage, evaluation of the interview and translation methods <u>Administrative Personal Data:</u> **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. **Benefit Recipient History (Leistungsempfängerhistorik - LeH):** Information on benefit receipt in accordance with Social Code Book III (SGB III) for recipients of unemployment benefit, unemployment assistance and maintenance allowance. **Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung - LHG):** Data on individuals receiving 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, authorized municipalities). **Jobseeker History (Arbeitssuchendenhistorik - ASU):** Information on job search activities that are recorded in BA procedures **Jobseeker History from XSozial-BA-SGB II (Arbeitsuchendenhistorik aus XSozial-BA-SGB II - XASU):** Information on job search activity reported to the BA by authorized municipalities via the transmission standard XSozial-BA-SGB II. **Participants-in-Measures History Files (Maßnahmeteilnahmehistorik - MTH):** Information on participation in employment and training measures (not including measures of authorized municipalities).

Categories	Notes
	Participants-in-Measures History File from XSozial-BA-SGB II (Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II - XMTH):
	Information on participation in employment and training measures reported to the BA by authorized municipalities via the transmission standard XSozial-BA-SGB II.
	Administrative Establishment Data
	Core dataset:
	Establishment variables (e.g. establishment ID, economic sector, federal state), employment structure of the employees (e.g. number of employees by sex and type of employment), age structure of the staff, structure of the staff by education and vocational training, structure of the staff by classification of occupation (Blossfeld classification), average wage of full-time employees.
	Extended Module – employee flows:
	Total number of entries, structure of entries by sex, type of occupation, classification of occupation (Blossfeld classification), re-employment, age; total number of exits, structure of exits by sex, type of occupation, classification of occupation (Blossfeld classification), seniority, age.
	Extended Module – establishment dynamics:
	Founding year, type of foundation, auxiliary variables to classify the foundation, year of closure, type of closure, auxiliary variables to classify the closure
Data unit(s)	Survey data:
	Anchor persons: Individuals who have arrived in Germany since 2013 and submitted their asylum application by the end of June 2016 at the latest.
	Family members: All close adult family members of the anchor persons, provided they live in the same household.
	Administrative data:
	Employees covered by social security (including marginal part-time employees from 1999 onwards), benefit recipients, jobseekers, participants in measures, establishments
Number of cases	Survey data:
	Wave 1: 4,465 individuals (in 3,289 households).
	• Of which 2,581 linked persons.
	Wave 2: 5,595 persons (in 3,822 households)
	• Of which 3,568 linked persons.
	Wave 3: 4,376 persons (in 3,061 households)
	• Of which 2,576 linked persons.
	Administrative individual data:
	4,549 persons with 121,509 observations
Period covered	Survey data:
	Wave 1: 2016
	Wave 2: 2017
	Wave 3: 2018
	Administrative data:
	The period covered depends on the data source.
	BeH: 01.01.1975 - 31.12.2018

Categories	Notes
	LeH: 01.01.1975 - 31.12. 2018
	ASU: 01.01.1997 - 31.12. 2018
	LHG: 01.01.2005 - 31.12. 2018
	XASU: 01.01.2005 - 31.12. 2018
	MTH: 01.01.2000 - 31.12. 2018
	XMTH: 01.01.2005 - 31.12. 2018
Regional structure	Survey data: German federal states (Bundesländer) Administrative data: German federal states (Bundesländer), districts (Kreise)
Survey design	Survey data: The sample was drawn from the Central Register of Foreigners (AZR) and is representative of asylum seekers who arrived between Jan. 1, 2013 and Jan. 31, 2016 and filed their asylum application by the end of June 2016 at the latest. In the 2017 survey, protection seekers who arrived by Dec. 31, 2016, and were registered by Jan. 1, 2017, were are also included. Asylum seekers with a higher likelihood of remaining in Germany at the time of sampling were over-sampled in the drawing due to the longitudinal design (e.g., Afghans, Iraqis, and Syrians). In addition, women and individuals over the age of 30 were also over-sampled. With appropriate weighting, it is nevertheless possible to make representative statements about refugees who entered Germany in the above-mentioned period. For more information on the survey design, see Brücker, Rothe, and Schupp (2017). Administrative data: Identification of respondents in IAB's Integrated Employment Biographies.
Institutions involved	Survey data: Cooperation between the Institute for Labour Market Research (IAB), the Research Data Centre of the Federal Office for Migration and Refugees (BAMF) and the Socio-Economic Panel at DIW Berlin. Implementation: Kantar Public (formerly TNS Infratest Sozialforschung) Administrative data: Social security agencies, Federal Employment Agency (Bundesagentur für Arbeit).
Frequency of data collection	Survey data: Yearly (Panel). Administrative data: Continuous.
Date of territorial allocation	Survey data: At the time of the survey. Administrative data: Territorial allocation updated as of 31.12.2018
File format	Stata
File size	Survey data: 552 MB (Stata)

Categories	Notes
	Administrative data:
	39 MB (Stata)
File organisation	Survey data:
	Personal, household and biographical data sets. The structure of the data sets correspond to the structure of the IAB-BAMF-SOEP survey of refugees.
	Administrative data:
	The data are stored in two files. One contains individual-level information and the other establishment-related information. Further files with additional information on establishments can be provided upon request.
Data access	Guest visit, remote data processing
Degree of anonymisation	Weakly anonymised
Sensitive variables	Survey data:
	None
	Individual administrative data:
	Month of birth (gebmon), Nationality (nation), occupational sub-group (beruf2010_4), place of residence: employment agency (wo_aa), place of resi-dence: district (Kreis) (wo_kreis)
	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 the data and data documentation	Data:
	"This article is based on data from the IAB-BAMF-SOEP survey of refugees (IAB-BAMF-SOEP) combined with administrative data from the IAB (1975-2018). The data was accessed via a research visit at the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research and/or via controlled remote data processing at the FDZ. DOI: 10.5164/IAB.IAB-BAMF-SOEP-ADIAB7518.de.en.v1"
	Data documentation:
	Keita, Sekou; Trübswetter, Parvati (2020): IAB-BAMF-SOEP Survey of Refugees linked to administrative data of the IAB. FDZ-Datenreport, 15/2020 (en), Nürnberg.
	DOI: 10.5164/IAB.FDZD.2015.en.v1
Dataset version	IAB-BAMF-SOEP-ADIAB 1975-2018 (IAB-BAMF-SOEP-ADIAB 7518); DOI: 10.5164/IAB.IAB-BAMF-SOEP-ADIAB7518.de.en.v1

Details on various access options, on the requirements for use and on applying for the dataset can be found on the FDZ website at https://fdz.iab.de/.

1.4   List of variables of the administrative data

The overview of variables in Table 4 lists the variable names and the longer descriptions of variables. It also provides an overview of whether and how well variables are filled in the individual data sources. Table 3 illustrates the meaning of the shading, which indicates the degree of completeness per variable and source in Table 4. Due to the high number of characteristics in the survey data, the IAB-BAMF-SOEP variables are only listed in the corresponding original documentation.

Table 3:   Degrees of completeness of variables

S	Variable is available for the data source. Degree of completeness > 0.85
G	Variable is available for the data source. Lower or varying degree of completeness, see description of variable and frequency count
W	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 4:   List of variables of the administrative data with degree of completeness

List of variables	BHP Basis	Page	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Identifiers		**39**				W			
Person number from the IAB-BAMF-SOEP Survey of Refugees (pid)		39	S	S	S	S	S	S	S
System-free person number (persnr)		39	S	S	S	S	S	S	S
Establishment ID (betnr)	*	40	S	W	W	W	W	W	W
Generated technical variables		**40**							
Counter per person (spell)		40	S	S	S	S	S	S	S
Source of spell (quelle)		41	S	S	S	S	S	S	S
Year (jahr)	*	41	S	W	W	W	W	W	W
Period of validity		**41**				W			
Original start date (begorig)		41				S			
Original end date (endorig)		42	S	S	S	S	S	S	S
Episode start date (begepi)		42	S	S	S	S	S	S	S
Episode end date (endepi)		42	S	S	S	S	S	S	S
Personal information		**43**	W	S	S	S	S	S	S
Gender (frau)		**43**				S			

List of variables	BHP Basis	Page	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Year of birth (gebjahr)		43	S	S	S	S	S	S	S
Month of birth (gebmon)		43	S	S	S	S	S	S	S
Nationality (nation)		43	S	S	S	S	S	S	S
Nationality, grouped (nation_gr)		44	S	S	S	S	S	S	S
Marital status (famst)		44	W	S	S	S	S	S	S
Number of children (kind)		44	W	S	S	S	S	G	G
Vocational training (ausbildung)		45	G	S	S	G	G	G	G
School leaving qualification (schule)		47	G	W	W	W	W	W	W
Information on employment, benefit receipt and job search		**48**	W	W	W	S	G	S	S
Daily wage/daily benefit (tentgelt)		48	S	S	S	S	S	S	S
Occupation - current/most recent (KldB 1988) (beruf)		49	S	S	W	W	W	W	W
Occupational group - current/most recent (KldB 2010), 3-digit (beruf2010_3)		50	S	W	W	G	W	G	G
Occupational sub-group - current/most recent (KldB 2010), 4-digit (beruf2010_4)		51	S	S	W	G	G	G	G
Level of requirement - current/most recent job (KldB 2010) (niveau)		51	S	S	W	G	G	G	G
Part-time (teilzeit)		52	S	S	W	G	G	G	G
Occupational status and working hours (stib)		52	S	W	W	W	W	W	W
Employment status (erwstat)		53	S	W	W	W	W	W	W
Transition zone (gleitz)		54	G	S	S	S	S	G	G
Temporary agency work (leih)		55	G	W	W	W	W	W	W
Fixed-term contract (befrist)		55	G	W	W	W	W	W	W
Reason of cancellation/notification/termination (grund)		55	S	W	W	W	W	W	W
Employment status prior to job search (estatvor)		56	W	S	G	S	G	W	W
Employment status after job search (estatnach)		56	W	W	W	S	G	W	W
Client profile (profil)		57	W	W	W	S	W	W	W

List of variables	BHP Basis	Page	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Reason for end of previous employment (art_kuend)		57	W	W	W	G	W	G	G
Working hours of job application (arbzeit)		57	W	W	W	G	W	G	G
Residual claim/planned duration (restanspruch)		58	W	W	W	G	W	G	G
Type of provider (traeger)		58	W	S	W	W	W	S	S
Start date of unemployment (alo_beg)		59	G	W	S	S	S	S	S
Duration of unemployment (alo_dau)		59	S	G	G	G	G	G	G
Location data		**59**	W	S	S	S	S	S	S
Place of residence - district (Kreis) (wo_kreis)		59	G						
Place of residence - federal state (Bundesland) (wo_bula)		60	G	G	S	S	S	S	S
Place of residence - employment agency (Arbeitsagentur) (wo_aa)		60	G	G	S	S	S	S	S
Place of residence - regional directorate (Regionaldirektion) (wo_rd)		61	G	G	S	S	S	S	S
Establishment variables		**61**	W	G	S	S	S	S	S
Classification of economic activities 73, groups (w73_3)	*	61	G						
Classification of economic activities 93, sub-classes (w93_5)	*	61	G	W	W	W	W	W	W
Classification of economic activities 93, groups (w93_3)	*	62	G	W	W	W	W	W	W
Classification of economic activities 03, sub-classes (w03_5)	*	62	G	W	W	W	W	W	W
Classification of economic activities 03, groups (w03_3)	*	63	G	W	W	W	W	W	W
Classification of economic activities 08, sub-classes (w08_5)	*	63	G	W	W	W	W	W	W
Classification of economic activities 08, groups (w08_3)	*	63	G	W	W	W	W	W	W

List of variables	BHP Basis	Page	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Classification of economic activities 73, completed by extrapolation/imputation, groups (w73_3_gen)	*	64	S	W	W	W	W	W	W
Classification of economic activities 73, type of imputation w73_3, groups (group_w73_3)	*	64	S	W	W	W	W	W	W
Classification of economic activities 93, completed by extrapolation/imputation, groups (w93_3_gen)	*	64	S	W	W	W	W	W	W
Classification of economic activities 93, type of imputation, groups (group_w93_3)	*	65	S	W	W	W	W	W	W
Classification of economic activities 08, completed by extrapolation/imputation, groups (w08_3_gen)	*	65	S	W	W	W	W	W	W
Classification of economic activities 08, type of imputation 08, groups (group_w08_3)	*	65	S	W	W	W	W	W	W
Year of first appearance (grd_jahr)	*	66	S	W	W	W	W	W	W
First appearance (grd_dat)	*	66	S	W	W	W	W	W	W
Year of last appearance (lzt_jahr)	*	66	S	W	W	W	W	W	W
Last appearance (lzt_dat)	*	67	S	W	W	W	W	W	W
Total number of employees (az_ges)	*	67	S	W	W	W	W	W	W
Number of full-time employees (regular workers + others) (az_vz)	*	67	S	W	W	W	W	W	W
Number of employees in marginal part-time employment (az_gf)	*	67	G	W	W	W	W	W	W
Mean imputed wage all full-time employees (te_imp_mw)	*	68	S	W	W	W	W	W	W
Place of work - district (Kreis) (ao_kreis)	*	68	S	W	W	W	W	W	W
Place of work - federal state (Bundesland) (ao_bula)	*	68	S	W	W	W	W	W	W

1.5 Volume structure and time periods

The number of person IDs in the IEB data is higher than the number of successfully linked data because some IEB accounts are split into multiple IEB accounts for administrative reasons between different versions. For example, 24 individuals have multiple IEB accounts in version 14 of the IEB, even though they were linked to a unique IEB account in version 13. The number of data lines in the administrative personal data is higher than the number of persons shown here, since there are overlaps between employment, benefit receipt periods, job search or participation in measures, which each receive their own data line. This also applies to the spell data recorded in the survey data. For some of the IAB-BAMF-SOEP respondents, no administrative data could be found in the IEB, although they were successfully linked. Table 5 therefore shows for how many IAB-BAMF-SOEP participants IEB data are available and how many rows they have in the administrative personal data.

Table 5:   Volume structure

IAB-BAMF-SOEP respondents	7.950
Respondents with available information for the RL	5.721
Successfully linked respondents	4.522
Person IDs in IEB data	4.549
Related rows in administrative personal data	121.509

Source: IAB-BAMF-SOEP-ADIAB, own calculations.

Since the individual data sets come from different sources, it is not possible for all partial data sets to cover the same period. The IAB-BAMF-SOEP survey data cover the observation periods of waves 1 to 3, which were collected in the years 2016 to 2018. Analogous to the data margins of the other administrative data products of the FDZ, the administrative personal and establishment data in IAB-BAMF-SOEP-ADIAB are available for the period 1975 - 2018. For the establishment data, the information refers to June 30 of the respective year. For individual persons, the periods may be significantly shorter, depending on their participation in surveys or their statuses relevant for the administrative data.

2 Data sources

2.1 IAB-BAMF-SOEP survey data

The aim of the IAB-BAMF-SOEP survey of refugees is to create a comprehensive and representative data basis on the fugitives living in Germany. The sample is representative of refugees who came to Germany between 2013 and 2017 and who have applied for asylum. The survey has been conducted annually since 2016 and covers the migration, education and employment biographies of the refugees, the reasons and routes of flight and collects personality traits, attitudes and information on the health status of the refugees. Other important parts of the study are accommodation, asylum procedures, integration into the German labor market, the educational system and other areas of society, as well as accompanying political measures. The individual questionnaires

comprise a total of about 450 questions. The head of household is asked a additional 100 questions on the current housing situation.

2.2   Administrative Data of the IEB

The administrative individual data were drawn from the Integrated Employment Biographies (IEB) of the IAB. They combine data from five different sources, each of which may contain information from different administrative procedures. Figure 2 illustrates the data records that are taken over from the IEB.

Figure 2:   Data sources of the IAB-BAMF-SOEP-ADIAB
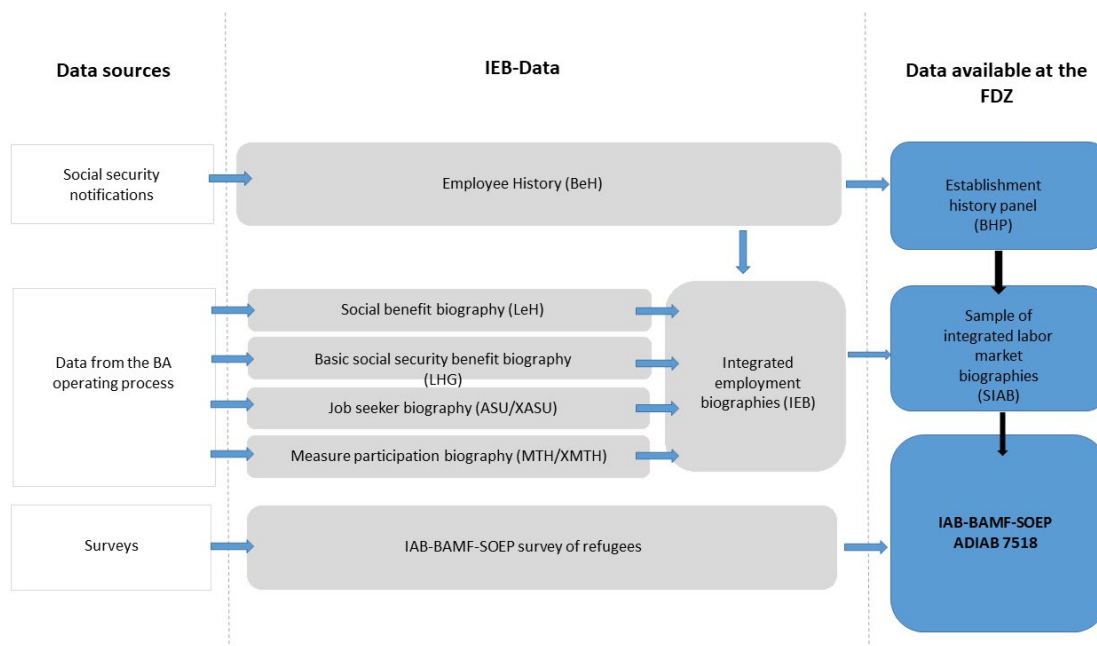
2.2.1   Employee History (BeH)

The source of data regarding employment is the Employee History (Beschäftigtenhistorik - BeH) of the IAB. The data basis is the integrated notification procedure for health, pension and unemployment insurance, which came into effect as of 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 students[1] (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

---

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

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.

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 reported for the 30th of June of a year. Unlike the data on individuals, the establishment variables are therefore not spell data but are only valid precisely for June 30 of each year[2].

Box 1: Example code for Stata 14

```
use IAB-BAMF-SOEP-ADIAB_7518_v1.dta
gen jahr = year(begepi)
sort betnr jahr
merge m:1 betnr jahr using IAB-BAMF-SOEP-ADIAB_7518_v1_bhp_ba-
sis_v1.dta
```

The Basis Establishment File is linked with the Individual File via the programme-specific commands of the software packages used for preparing and analysing the data. In Stata, for instance, the two files can be linked using the "merge" command in connection with the relevant paths (see Box 1).

2.2.2   Benefit Recipient History (LeH)

The Benefit Recipient History (Leistungsempfänger-Historik - LeH) of the IAB covers periods during which individuals receive earnings replacement benefits from the Federal Employment Agency (sphere of Social Code Book III). The benefits comprise unemployment benefit, unemployment assistance, maintenance allowance, and contributions paid by the BA to private health or care insurance while benefits are being drawn. Benefits in the context of Social Code Book II (e.g., unemployment benefit II) are not included in the data. 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.

2.2.3   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

---

[2] 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.

individuals associated with the benefit unit. In the IAB-BAMF-SOEP-ADIAB, 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 IAB-BAMF-SOEP-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,[3]

- 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.[4] 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 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).

2.2.4   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

---

[3] 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.

[4] 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.

standard. The earliest available data in the ASU are from 1 January 1997 and in the XASU from 1 January 2005, respectively.

2.2.5   Participants-In-Measures History Files (MTH/XMTH)

The Participants-In-Measures History Files (Maßnahmeteilnahmehistoriken - MTH) contain information that can be assigned to different legal spheres. First, they contain active labour market policy measures in accordance with Social Code Book III and participation in such measures. Second, the MTH contain measures in the legal sphere of Social Code Book II if these are recorded in BA administrative procedures. Measures implemented by authorised municipalities or opting local authorities are recorded in the XMTH from 2005 onwards. The earliest available data in the MTH are from 1 January 2000, those in the XMTH are from 1 January 2005.

3   Data preparation and sampling procedure

3.1   IAB-BAMF-SOEP survey data

The sampling concept of the IAB-BAMF-SOEP study is presented by Brücker, Rothe, and Schupp (2017). The FDZ Methods Report 02/2019 (Kosyakova et al., 2019) contains detailed information on data collection and processing. The household and personal questionnaires used, as well as the documentation of the survey data used for the linked data set in the format of the Scientific Use File, also offered by the FDZ, are available at https://fdz.iab.de/en/FDZ_Individual_Data/iab-bamf-soep/IAB-BAMF-SOEP-SUF1618v1.aspx. Further publications on methodological aspects of the IAB-BAMF-SOEP study are also available.

3.2   Administrative data from the IEB

3.2.1   Data corrections

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

- Observations in which the age is under 13 or over 75 are deleted.
- Observations whose end date precedes the start date are deleted.
- Inconsistent information on gender or date of birth within an account is corrected.
- 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.

3.2.1.1 Employee History (BeH)

- To capture a person group that is as constant as possible over time, some person groups for which data are not available throughout the entire observation period are excluded. From the reporting year 2011 onwards, the BeH data originate from newly designed data sources. 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. The person groups 123, 124, and 127 were 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 new person groups 121 and 122.

- 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.

3.2.1.2 Benefit Recipient History (LeH)

- Observations without a valid start date are excluded.

- If the end date for the receipt of unemployment assistance precedes the start date by one day and the spell was not deleted, then the end date is increased by one year.

- Between 2004 and 2006 the notification procedure from which the data originate was changed. Overlaps occurring between the old and the new procedures were corrected.

- Observations with no end date or an invalid end date are excluded, 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 in the BeH.

3.2.1.3 Unemployment Benefit II Recipient History (LHG)

- Observations without a BA client number are deleted.

- Cancelled data records are not used.

- Only observations of people who are capable of work and people under the age of 65 are included.

- In each case, non-overlapping periods of benefit entitlement of a person in a certain benefit unit (BG) are created. New observations can start 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 3.2.1.7),
- 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 April 1st, 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 reasons 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 4.2.2) were corrected as far as possible.

- The territorial allocations are corrected in the same way as in the BeH.

3.2.1.4 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 in the BeH.

3.2.1.5 Participants-In-Measures History File (MTH)

- Observations generated more than a year after the end of the measure are deleted if another observation exists that was generated within the year after completion of the measure.

- Only the most recent record of an individual case of participation in a measure is used.

- Only cases of participation in measures that are classified as "actually took place" are included in the IEB. Cases of participation that did not take place or have not yet taken place are deleted. Cases of participation are also classed as not having taken place when a deletion date is set during the participation in a measure.

- Certain types of measure are not included. These include services to support careers advice and job placement, mobility assistance and pure rehabilitation measures.

- The territorial allocations are corrected in the same way as in the BeH.

3.2.1.6 Participants-In-Measures History File XSozial-BA-SGB II (XMTH)

- For XMTH, the particular challenge is to identify and handle multiple notifications of participation that are caused by technical or organisational issues. The causes of these duplicate notifications are, on the one hand, a missing identification number for participation in measures until April 2009 or improper handling of the subsequently introduced promotion ID and, on the other hand, a change of provider numbers in connection with mergers or other reorganisations of institutions responsible for implementing SGB II due to the reorganisation of the employment agencies in 2012/2013. In addition, provider-specific problems arose with version or product changes of the municipal software or with version changes of the XSozial standard. At times, this results in a considerable number of reports of participations in measures of the same type, which partly or completely overlap in time per participant. Real duplicates are identified and sorted out because they are highly likely to be technical duplicates. Hidden duplicates can also be identified and sorted out. In principle, the rule always applies that the most up-to-date information is retained.

- In addition, notifications with a starting date before 2005 are excluded.

- Overlapping and immediately adjacent notifications of the same type of measure (from XSozial) are combined to one observation. Measure-specific characteristics of the combined periods are set to system missings.

- Subsequently, the remaining characteristics are compiled and calculated. For this, consolidated individual data from XSozial-histories and LHG are used.

- Finally, the following measures are excluded:
  – one-off benefits (like UBV/Mobi-/Vermittlungsbudget /LES) and
  – specific rehabilitation measures.

  The territorial allocations are corrected in the same way as in the BeH.

3.2.1.7 SGB II anonymisation

In order to reduce the risk of de-anonymization, only the year of birth is available in the IAB-BAMF-SOEP-ADIAB by default. 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, such as the first day of retirement and/or the day before.

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".

3.2.2   Episode splitting

The administrative individual data are available with "split" episodes. If observations overlap within an account, these observations are replaced by artificial observations with new dates so that completely parallel periods and non-overlapping periods (episode splitting) 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`).

It is advisable to sort parallel observations generated by the splitting procedure in a consistent manner.

Figure 3:   Episode splitting

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

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

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

3.2.3 Missing values

In the administrative data missing values are coded as follows:

Table 6: Coding of 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 Table 4) or is not available for a certain period.

3.3 Data Linkage

The starting point for linking the survey and process data are the respondents of the IAB-BAMF-SOEP survey of refugees. During the interview, they were asked for their consent to link their survey data with the administrative data available at the IAB for research purposes. The questionnaire section with the declaration of consent reads as follows:

„I agree that the data held at the Institut für Arbeitsmarkt-und Berufsforschung (IAB) in Nuremberg can be included in the study, merged exclusively with my survey data from the LIVING IN GERMANY study, and stored anonymously. The study will take place in such a way that it will not be possible to associate the data with my identity. I have been informed that all data protection regulations will be observed. I reserve the right to withdraw my consent at any time.“

Respondents had to agree to the above question in order to become part of the sample for the linkage. For the linkage procedure, the respondents can be divided into two groups. Anchor persons are those who were included in the sample by drawing from the Integrated Employment Biographies and thus have an indicator with which they can be traced back into the IEB. Family members of the anchor persons were also interviewed, but do not have an indicator with which the data can be linked with the IEB. These respondents have to be searched for in the IEB with the help of linkage procedures.

In waves 1 to 3 of the IAB-BAMF-SOEP sample of refugees, a total of 7,950 persons were interviewed at least once, of whom 6,670 (around 84 per cent) gave their consent to linkage. The BAMF provided the IAB with information on 5,721 persons who could be searched for in the IEB. For 4,522 persons, the survey data could be assigned to an account (in some cases two accounts) in the administrative data.

After completion of the survey, an attempt was made to link respondents who were ready to participate. Generated technical features in the data set indicate for each respondent which linking method was used to achieve a successful link. This allows each user to decide at their own discretion which of the cases should be used in their analyses.

3.3.1 Cleaning and standardization of the linkage variables

The following personal characteristics from all available address data sources were used to search for the family members:

- First and last name
- Gender

- Address (postal code, town, street, house number)
- Date of birth

Before linkage, the characteristics were subjected to cleaning routines in order to achieve as high a match as possible between the two data sources.

In concrete terms, German umlauts and special characters were replaced by their respective ASCII-compliant equivalents, all letters were converted into capital letters and unnecessary blanks and punctuation marks were removed. Several first or last names were combined in one field without spaces. In addition to these basic clean-ups, variable-specific routines were carried out to standardise the data. For example, the house number was extracted from a combined street/house number field and all other address additions (e.g. floor, flat no.) were deleted. For street names, frequently occurring patterns and abbreviations were standardised. For place names, appendices are standardised (e.g. for places with additional information such as river or region) or removed (e.g. for Berlin-Kreuzberg).

3.3.2 Linkage of anchor persons

First, the anchor persons were linked to the IEB data. Since the data on the anchor persons already contain the identifiers from the IEB data due to the type of drawing, they can be directly merged. Due to the recoding between IEB versions, it can happen that several IEB person numbers are also available for anchor persons. Furthermore, quality assurance of the linkage also makes sense for anchor persons, since it is possible that the interviewer has determined the wrong person as the anchor person of the household.

3.3.3 Exact record linkage

For the family members, an exact matching of the fields between the entries of the two data sources for the family members was first carried out.

In the first step, the complete match in all available linkage-relevant fields was required, i.e. first and last name, date of birth, gender, street, house number, postcode and town. In the next two steps, a deviation was tolerated once in the postcode field and once in the house number field for persons who were not linked in the first step, while the other fields had to match exactly. The generated technical characteristic match_type shows in which step the linkage took place. Only characteristic combinations were considered in which the risk of linking two different persons (so-called false-positive matches) was very low. All in all, the following combinations of characteristics were used in the exact record linkage to check for a match between the two data sources:

1. surname, first name, date of birth, sex, postcode, street, house number, town (complete)

2. surname, first name, date of birth, sex, street, house number, town

3. surname, first name, date of birth, sex, postcode, street, town

4. surname, first name, date of birth, sex, street, town

3.3.4 Probabilistic record linkage

For persons for whom no match could be found in the administrative data through any of the previous steps, error-tolerant matching procedures were then used (cf. Herzog et al., 2007). A similarity measure was calculated for each pair of variables to be matched from both data sources and

these were summed up in a summary quality index. The calculations were carried out with the software Merge Tool-Box (MTB, version 0.742, cf. Schnell et al. (2005)) and with Stata. In the calculation of the similarity measure, probabilistic matching was applied, which takes into account that the probability of a match of characteristics depends on their variance in the population (see Fellegi/Sunter, 1969; Christen, 2012, pp. 133-137). Here, so-called m- and u-parameters are calculated, which indicate the probability with which the information on a person varies between two different sources or how likely the occurrence of an identical characteristic is with different persons. For example, a match for the variable gender is less indicative of an actual match than a match for the surname and is therefore weighted lower. All available characteristics are compared between the two data sources and an additive similarity measure is calculated. The higher the value of this index, the greater the overall similarity of the two rows and the more likely there is an actual match.

The address pairs were sorted according to the similarity measure and subjected to a visual comparison. The resulting links were classified as probabilistic matches and included in the data set.

4 Data quality and problems

4.1 IAB-BAMF-SOEP survey data

Non-response, weighting and other aspects of data quality of the IAB-BAMF-SOEP data are discussed in Brücker, Rothe, and Schupp (2017). The FDZ Methods Report 02/2019 (Kosyakova et al., 2019) examines methods to identify false interviews. At https://fdz.iab.de/en/FDZ_Individual_Data/iab-bamf-soep/IAB-BAMF-SOEP-SUF1618v1.aspx there are numerous other methodological publications dealing with the quality of the IAB-BAMF-SOEP data.

4.2 Complete IEB

4.2.1 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 7 makes no claims to be exhaustive.

Table 7:   Biographical gaps and possible ways of identifying them

Biographical gab	Information on gap, identifiable using the details in the "grund" variable in the preceding observation of the source, if necessary
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 3.2.1.4)	BeH, LeH, ASU
Persons receiving old-age pension without employment if not a member of a benefit unit	LeH, LHG, ASU
Individuals on maternity leave / parental leave	XASU
Recipients of early retirement benefits	LeH, ASU
Trade professionals working from home	
Employees working short time	ASU
Persons in youth welfare facilities, in vocational training centres, approved workshops or similar facilities for disabled persons	ASU
Participants in programmes to support participation in working life (people in rehabilitation)	ASU
(Sideline) farmers	
Caregivers according to Section 19 SBG XI	
Conscripts	BeH, LeH, LHG, ASU, XASU
Persons in reserve duty training	BeH, LeH, LHG, ASU, XASU
Persons fulfilling community service	BeH, LeH, LHG, ASU, XASU
Persons fulfilling a voluntary social or ecological year instead of community service	
Other people not registered with the statutory pension insurance or the Federal Employment Agency (e.g., sabbatical, funding from personal assets or pensions, emigration, employment abroad, voluntary work etc.)	BeH, LeH, ASU
Strikers in cases where the strike lasts more than a month	LeH
Social assistance recipients (prior to the introduction of SGB II in 2005), recipients of welfare payments (according to SGB II)	
SGB-II recipients whose providers have experienced delivery failures	
Recipients of compensation according to FELEG (Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit, Act on Support in Case of Termination of Farming Activities)	

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

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

4.3   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 2.2) 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[5].

- The variable "Occupational status and working hours" (stib) is only filled for reports that were submitted before the introduction of the new occupation code, and the RDC does not extrapolate or impute for later reports. The categories of the variable stib that can be consistently observed over the whole observation period (occupational status as a trainee, distinction between part-time and full-time) are to be reproduced in the variables erwstat and teilzeit.

- The introduction of the new occupation code in 2011 led to a number of problems. For example, during the transition period granted to employers in the social security notification procedure,[6] there was a temporary increase in the number of missing details. Analyses of the BA statistics (Bertat et al., 2013, p. 10) show that in 20 to 30 percent of cases no information was available in the new or converted variables "occupation/activity performed", "working time" and "vocational education and training" after the switch. This situation began to improve significantly in the first half of 2013. In order to improve the quality of the "working time" variable in the transition period, Ludsteck/Thomsen (2016) developed an imputation procedure to replace the missing values by imputed values. The imputed data are included in the IAB-BAMF-SOEP_ADIAB 7518. 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

---

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

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

(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 present version of the IEB the year 2014 is the last year with a degree of completeness of BeH observations of 100%.[7] 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, and that those for 2017 are underreported to a 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. This means that in individual cases the establishment data, e.g. establishment size, are grossly incorrect and will change considerably in subsequent versions.

- In 1984, a change was made in the employment notification procedure. From that time onwards, one-off payments of gross earned income were reported as part of the annual earnings subject to social security contributions, which leads to an increase in the average daily wage. In particular, the proportion of wages and salaries above the upper earnings limit increased considerably from that year onwards (see Bender et al. 1996).

- For the years 1992 until 2000, noticeable decreases and increases in the number of notifications were observed. Decreases can be observed especially for the following 10 districts: Braunschweig (03101), Wolfsburg (03103), Emden (03402), Kassel (06633), Essen (05113), Neuss (05162), Erftkreis (05362), Hersfeld-Rotenburg (06632), Miltenberg (09676) and Kempten (Allgäu) (09763). This is due to notification problems of one or more establishments in these regions.

- Considerable decreases were also observed for the districts Salzgitter (03102) and Hoyerswerda (14264).

- Concerning the notifications for full-time employment, especially the districts Main-Taunus (06436) and Alzey-Worms (07331) are noteworthy. They feature above-average rises. Also in this case, the reasons are notification problems at one or more establishments in these regions.

- In the years 1996 to 1998, the values 841-844 (doctors and pharmacies) within the 'occupation – activity performed' 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

---

[7] Due to a redesign of the data basis of the BeH, at the time of the preparation of the BeH as an exception only the 30-month files was 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.

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.

4.4   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 an errorin 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 the 1st of July, 2004, periods of exclusion from benefits and of benefit suspensions 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.

4.5   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 exclusively 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 due to changes in the type of intitution 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 centers, 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 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 IAB-BAMF-SOEP-ADIAB.

- 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.

4.6   Jobseeker Histories (ASU/XASU)

4.6.1   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 ARGEn, 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.

4.6.2 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.

4.7   Participants-In-Measures History Files (MTH/XMTH)

4.7.1   MTH

- The MTH is incomplete for measures with a start date before 1 January 2000.

- As of 1 January 2005, there is an inconsistency in the data as participants in measures were allocated to different institutions with the introduction of Social Code Book II (see Section 2.2.5 and 4.2.2).

- The MTH only contains notifications that are recorded in BA procedures. The use of these procedures in cooperations between employment agencies and municipalities/separated responsibilities/municipalities exercising their duties separately increases continuously between 2005 and 2007. The notifications for these institutions are complete from March 2007 onwards. Measures that are reported by authorised municipalities via the XSozial standard are contained in the XMTH.

- Because of the reorganisation of the institutions responsible for implementing SGB-II in 2011 to 2014 (see Section 4.2.2), a split of the documentation of participations in measures in the MTH and in the XMTH may occur when there is a change in the reporting procedure [8]. This might result in a split or a duplication of the spell of the measure (see Section 3.2.1.5).

- In the case of notifications regarding the bridging allowance (*Überbrückungsgeld*) the maximum permissible duration of six months is sometimes exceeded. In most cases, this can be explained by a default setting in the input mask of the data recording system.

- The MTH is supplemented by applicant characteristics (e.g., vocational training) from other BA procedures. For these variables the administrative procedure was switched from coArb to VerBIS in 2006. The same quality limitations as for the ASU apply here.

4.7.2   XMTH

- In the years 2005 to 2007 the notifications of participation in measures are incomplete. The degree of under-reporting is unclear due to a lack of or inaccurate comparative figures.

- Between 2005 and the beginning of 2017, about 13% of all institutions responsible for implementing SGB II report almost no municipal integration benefits (formerly accompanying benefits), which are included in the summarised category "other support". Many other job centers report only temporarily and/or only selected types. This leads to an under-recording of the "other support" category.

---

[8] Further information concerning the territory structure of the institutions responsible for implementing Social Code Book II and relevant changes is available at https://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikationen/Regionale-Gliederungen/Gebietsstruktur-Traeger-Grundsicherung-Nav.html

- Over-registration of participations: Total stocks are considered to be stable from the end of 2008. However, the majority of the providers still register notification profiles which are temporarily conspicuous, and which raise doubts about the quality. Only the introduction of the XSozial promotion-ID in 2009 and the meanwhile several years of experience of all participants stabilise the reporting process and the subsequent data processing. Also, the new zkT, introduced in 2012, report inconspicuously for the most part, so that the scope and duration of presumed over-recording decrease noticeably.

- The reorganisation of SGB II providers in 2011-2014 (see Section 4.2.2) lead to a split in the documentation of participation in measures in MTH and XMTH if the reporting procedure was changed. This could result in a split of the measure spell, but also in duplications (see Section 3.2.1.6).

- The figures or persons received in the XMTH from monthly key date counts differ in several respects from 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 the IAB-BAMF-SOEP-ADIAB files under http://fdz.iab.de/en.aspx.

5.1 Identifiers

5.1.1 Person number from the IAB-BAMF-SOEP Survey of Refugees (pid)

Variable label	Person number from the IAB-BAMF-SOEP Survey of Refugees
Variable name	pid
Category	identifiers
Origin	IAB-BAMF-SOEP Survey of Refugees
Data type	numerical
Detailed description	The person number indicates which records belong to the same person in the survey data. It is not possible to infer characteristics of the person or original identifiers in the administrative data from the person number.

5.1.2 System-free person number (persnr)

Variable label	Self-generated identification of IEB accounts
Variable name	persnr
Category	identifiers
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	The account name indicates which records belong to the same IEB person account. It is not possible to infer the characteristics of the person or original identifiers in the administrative data from the account name.
	As there is no uniform person identifier in the different data sources, the allocation of information from the different data sources (e.g. employment and benefit) to persons (accounts) is not always clear. In such cases, implausible employment histories may result.
	The creation of the cross-source person identifier is based on a heuristic developed by the BA.

5.1.3 Establishment ID (betnr)

Variable label	establishment ID
Variable name	betnr
Category	identifiers
Origin	BeH
Data type	numerical
Detailed description	The establishment ID indicates which observations belong to the same establishment. It is based on the establishment number allocated by the BA, which was replaced by an artificial number (further information on the allocation of establishment numbers by the BA can be found in Bender et al. (1996: p. 15 et seq. and pp. 27-30) as well as directly on the website of the establishment number service of the BA at https://www.arbeitsagentur.de/betriebsnummern-service/alles-wichtige. The establishment number and year specification can be used to merge individual and establishment information. For the establishment number, the following should be observed in general: 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. 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 industry. If they do **not** belong to the same industry, each branch office is regarded as a separate establishment and is given its own establishment number. 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. 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: 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. A **workplace** is a unit in which employees work and which is not allocated an establishment number according to the above-mentioned principles. A **company** as a term combines establishment premises and workplaces belonging to the same employer. 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. Establishment and establishment premises are synonyms; branch office is a synonym for subsidiary, district office, out-sourced office, workplace etc. if it is not an establishment.
Notes on quality	The establishment ID is only missing in a very small number of cases. These observations are notifications for the person group "205" (earnings notifications for casual workers). As establishment variables (place of work, economic activity, establishment size etc.) are merged via the establishment ID, they are missing in these observations.

5.2 Generated technical variables

5.2.1 Counter per person (spell)

Variable label	counter per person
Variable name	spell
Category	generated technical variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
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. Within employment notifications, persons subject to social insurance contributions are sorted before marginal employment notifications and higher daily wages before lower ones. One exception is one-time payments, which are sorted backwards.

5.2.2   Source of spell (quelle)

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

5.2.3   Year (jahr)

Variable label	year
Variable name	jahr
Category	generated technical variables
Origin	BeH
Data type	numerical
Detailed description	This variable 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 in Section 2.2.1 for an example code with the „merge"-command in Stata 14.

5.3   Period of validity

5.3.1   Original start date (begorig)

Variable label	original start date
Variable name	begorig
Category	period of validity
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	date
Detailed description	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 3.2.2). 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 3.2.1.7. **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. **2) LHG, ASU, XASU** Certain changes lead to the creation of a new observation (see Sections 3.2.1.3 and 3.2.1.4). begorig indicates the start date of the new period **3) MTH, XMTH** In MTH and XMTH, new data records are created when a change of provider takes place during participation in a labour market policy measure.

5.3.2   Original end date (endorig)

Variable label	original end date
Variable name	endorig
Category	period of validity
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	date
Detailed description	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 3.2.2). 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 3.2.1.7. **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. **2) LHG, ASU, XASU** Certain changes lead to the creation of a new observation (see Sections 3.2.1.3 and 3.2.1.4). endorig indicates the end date of the new period. **3) MTH, XMTH** In MTH and XMTH, new data records are created when a change of provider takes place during participation in a labour market policy measure.

5.3.3   Episode start date (begepi)

Variable label	episode start date
Variable name	begepi
Category	generated period of validity
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
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 3.2.2).

5.3.4   Episode end date (endepi)

Variable label	episode end date
Variable name	endepi
Category	generated period of validity
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
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 3.2.2).

5.4 Personal information

5.4.1 Gender (frau)

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

5.4.2 Year of birth (gebjahr)

Variable label	year of birth
Variable name	gebjahr
Category	personal variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	The year of birth is constant within one individual account.
Notes on quality	In the original data, it may happen that the date of birth changes between the data sources. This is corrected during the data preparation process. The information from the social security number is given highest priority here.

5.4.3 Month of birth (gebmon)

Variable label	month of birth
Variable name	gebmon
Category	personal variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical

Detailed description	The month of birth is constant within one individual account.
	One can use the variables Year of birth (gebjahr) and Month of birth (gebmon) to generate a variable in the date format JJJJmM (e.g., 1984m6) with the following syntax in Stata:
	`gen int gebdat = ym(gebjahr, gebmon)`
	`format gebdat %tm`

Notes on quality	In the original data, it may happen that the date of birth changes between the data sources. This is corrected during the data preparation process. The information from the social security number is given highest priority here.
Anonymisation	Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (gebjahr) is provided.

5.4.4 Nationality (nation)

Variable label	nationality
Variable name	nation
Category	personal variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
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.
Besonderheiten	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.

5.4.5 Nationality, grouped (nation_gr)

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

5.4.6 Marital status (famst)

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

5.4.7 Number of children (kind)

Variable label	number of children
Variable name	kind
Category	personal variables
Origin	LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical

Detailed description	This variable has a different meaning depending on the data source. **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 **2) ASU, MTH** In these sources, the value of the variable corresponds to the actual number of children under 15 living in the household. **3) LHG, XASU, XMTH** 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.
Notes on quality	**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. **2) ASU, MTH** Until the 30th of June 2006, only up to nine children could be recorded. The value zero does not exist. For observations prior to the 30th of 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.

5.4.8 Vocational training (ausbildung)

Variable label	vocational training
Variable name	ausbildung
Category	personal variables
Origin	BeH, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	It must be taken into account that this variable has a different meaning depending on the data source: **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 12 University without further specifications In notifications that rely on the new occupation code (see Section 4.3) 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. **2) ASU, MTH** 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: 1 Without vocational training 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 and MTH which have a start date from 1 July 2006 onwards, the following differentiated categories are available:

7	Voc. training not accepted in Germany
8	University degree not accepted in Germany
9	In-company voc. training/traineeship/in-school voc. training
10	Other exams
11	University of applied sciences without further specifications
12	University without further specifications
13	Doctorate
14	Bachelor (BA)
15	Bachelor (FH)
16	Bachelor (University)
17	Master (FH)
18	Master (University)
19	Diploma (BA)
20	Diploma (FH)
21	Diploma (University)
23	Undergraduate studies
24	Secondary/additional studies

3) XASU, XMTH

For spells that originate from these sources, the vocational education completed most recently is reported. The following values exist:

1	Without vocational training
2	In-company voc. training/traineeship/external voc. training
3	Technical school (voc. training)
4	Technical school (advanced voc. training)
5	University of applied sciences (FH)
6	University
7	Voc. training not accepted in Germany
8	University degree not accepted in Germany

For analyses that cover a longer period of time, the values can be aggregated as follows:

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	**1) BeH**
	"Changes in the vocational training status frequently occur at the same time as a change of establishment. This is because the notification data are compiled anew in the new firm. If, for example, an employee obtained a higher qualification via a part-time further training course while still working then this change of status is probably not recorded until he/she joins a new firm. It can generally be assumed that when a person is employed in a firm for a longer period, the personal data that they reported when they joined the firm is simply carried forward" (own translation of Meinken / Koch 2004, p. 63).
	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%.
	Missing values occur particularly frequently in the following groups: marginal part-time employees, part-time workers, foreign employees and workers from Eastern German. The reason for this is that the variable is not of particular importance as regards social security contributions (see Meinken/Koch, 2004, p. 63).
	2) ASU, MTH
	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 and MTH data sources 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. In the source MTH, older categories partially still occur even after 1 July 2006.
	3) XASU, XMTH
	The degree of completeness in the XASU and the XMTH is generally low.

5.4.9 School leaving qualification (schule)

Variable label	school leaving qualification
Variable name	schule
Category	personal variables
Origin	BeH, ASU, XASU, MTH, XMTH
Data type	numerical

	This variable contains the school leaving qualification. Different values are possible depending on the source.
	1) BeH With the switch to the new occupation code (see Section 4.30) the possible values of the variable change.
	The values from the old occupation code are: 5 Grade-/lower school certificate, intermediate school or equivalent qualification 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 9 Upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling
Detailed description	With the new occupation code the values are: 1 No school leaving certificate 4 Lower secondary school certificate/ grade school certificate 6 Intermediate school leaving certificate 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
	2) ASU, XASU, MTH, XMTH The following values are possible for these data sources: 1 No school leaving certificate 4 Lower secondary school certificate/ grade school certificate 6 Intermediate school leaving certificate 7 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college 9 Upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling
	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.
Notes on quality	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 and the XMTH it has been increasing continuously and has levelled off at over 2/3 since 2013 (XASU) and 2012 (XMTH), respectively. In the ASU and the MTH the degree of completeness is generally high.

5.5 Information on employment, benefit receipt and job search

5.5.1 Daily wage/daily benefit (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

	1) BeH
	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.
	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.
	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, employment interruption or end of employment notifications. In this case, it is irrelevant whether the upper earnings limit in the statutory pension insurance which is decisive for the notification period is exceeded as a result of this addition. However, such earnings notifications could also be due to incorrect details in the employment period. (The earnings in-formation, however, may be considered less error-prone due to its insurance relevance.)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.
Detailed description	A daily wage reported as 0 euros can be put down to "employment interruption notifications". During these periods, the employment relation-ship 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.
	From 2013 onwards, the number of notifications with a reason for deregistration of 54 (notification of a one-off wage) increases sharply (see Section 5.5.12). It is likely that special payments which were reported with the annual declarations before 2013 are now reported separately. It is therefore advisable to add these variable one-time payments to the corresponding wages for simultaneous employment episodes within the same establishment when analysing wages over time.
	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.
	2) LeH
	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.
	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.

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

Variable label	occupation – current/most recent (KldB 1988)
Variable name	beruf
Category	information on employment, benefit receipt and job search
Origin	BeH, ASU, XASU, MTH
Data type	numerical
Detailed description	**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 applies to one employee, the employer is required to select the job title that best defines the main activity performed (see BA 2005, p. V). For this, the employer encodes the employee's job in accordance with the "Classification of Occupations. Systematic and Alphabetical Directory of Job Titles" (published by the Federal Employment

Agency, Nuremberg, 1988), which contains approx. 25,000 job titles. The occupational classification consists of a 3-digit code and comprises about 330 values.

Employment notifications with an end date later than 30 November 2011 are reported using the new occupation code 2010 (KldB2010) (see Section 4.3). These values are transcoded to the KldB1988 via a priority switch. This results in inaccuracies.

2) ASU, XASU, MTH

The variable contains the occupation of the last job. See 1) with regard to the occupation code.

Notes on quality	**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. **2) ASU, MTH** The degree of completeness decreases significantly from 2014 to 2017. **3) XASU** The occupation variable is not filled for almost the entire period available.

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

Variable label	occupational group - current/most recent (KldB 2010), 3-digit
Variable name	beruf2010_3
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	**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 4.3). 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. **2) LeH, ASU, XASU, MTH, XMTH** The variable contains the occupation of the last job. See 1) with regard to the occupation code.
Notes on quality	**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. **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. **3) XASU, XMTH** The variable is not filled for almost the entire period available.

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

Variable label	occupational sub-group - current/most recent (KldB 2010), 4-digit
Variable name	beruf2010_4
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	**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 varia-ble '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 4.3). 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. **2) LeH, ASU, XASU, MTH, XMTH** The variable contains the occupation of the last job. See 1) with regard to the occupation code.
Notes on quality	**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. **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. **3) XASU, XMTH** 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.

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

Variable label	level of requirement - current/most recent (KldB 2010)
Variable name	niveau
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	**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 0). 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.

2) LeH, ASU, XASU, MTH, XMTH

The variable contains the occupation of the last job. See 1) with regard to the occupation code.

Notes on quality	**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. **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. **3) XASU, XMTH** The variable is not filled for almost the entire period available

5.5.6 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.
Notes on quality	There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. In order to reduce this problem, the working hours were imputed at the IAB for the period in question. Further information about the procedure can be found in Ludsteck/Thomsen (2016). A similar accumulation of missing values occurs in 1991 due to the reunification. No imputation is performed here.

5.5.7 Occupational status and working hours (stib)

Variable label	occupational status and working hours
Variable name	stib
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical

Detailed description	The employee's occupational status during the notification period is reported by the employer as part of the "employment details". The variable "occupational status and working hours" distinguishes between full-time and part-time employees. The decisive factor for this differentiation is the ratio between the contracted hours and the usual working hours in the establishment. For part-time employees the variable only indicates whether their working hours exceed a certain threshold. Until 1978, this threshold was 20 hours of work per week, between 1979 and 1987 it was 15 hours per week and since 1988 it is 18 hours per week.
	The variable only provides information regarding the occupational status for full-time employees, distinguishing among other things between blue-collar and white-collar employees in full-time employment and apprentices. The distinction between (full-time) blue-collar and white-collar employees is solely based on the type of compulsory pension insurance (Federal Social Insurance Office for Salaried Employees – BfA – for white collar workers, and State Social Insurance Office – LVA – for blue-collar workers). Master craftsmen and foremen are only included in a separate category if they are compulsorily insured in the workers' pension insurance. The assignment of a master craftsman or foreman to the blue-collar or white-collar employees can only be made via the respective pension provider. The "employees in vocational training" category covers not only apprentices, volunteers and interns but also semi-skilled trainees, students at colleges for health occupations and participants in subsidised further vocational training, retraining and induction training.
	If more than one code is eligible for an employee, the employer is required to classify the job according to the activity which is predominantly performed. If this cannot be determined clearly, the code of the higher occupational status is to be entered (see BA 2005, p. VI).
	Owing to the introduction of the new occupation code (see Section 4.3), however, this distinction is no longer possible. The variable „stib" is therefore only filled for notifications which date back to before the introduction of the new occupation code.
Notes on quality	There is a considerable number of missing values in 1991 due to the German reunification.

5.5.8 Employment status (erwstat)

Variable label	employment status
Variable name	erwstat
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	This variable takes on different values with different meanings for each data source.
	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.
	2) LeH
	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.

3) LHG

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.

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 3.2.1.7.

4) ASU, XASU

For ASU/XASU observations, the 'employment status' variable reports the job search status.

A distinction is made between those who are "employed 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.

"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.

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 3.2.1.7.

5) MTH, XMTH

For observations in the MTH and the XMTH, the 'employment status' variable indicates the measure-type category. This is the highest level in the hierarchy of the measure-type classifications of the BA.

Notes on quality	**1) LHG**
	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.
	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.
	2) ASU, XASU
	The categories "seeking advice" and "not seeking work" have only existed since the introduction of VerBIS (see Section 4.6). 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.

5.5.9 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.

5.5.10   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%.

5.5.11   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%.

5.5.12   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** 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. 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). **2) LeH** 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. **3) LHG** 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.

4) ASU

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'.

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):

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	

5) XASU

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.

Notes on quality	**1) BeH** From 2013 onwards, the number of notifications with a reason for deregistration of 54 (notification of a one-off payment) increases sharply. As long as an employment relationship exists, special payments that are paid out by March of the following year can be included into the usual notifications (mostly annual notifications) for the previous year. A separate notification with a reason for deregistration of 54 is then not required. Until 2012, the annual notifications could be submitted until mid-April; since 2013, they must now be submitted by mid-February at the latest. Special payments made in February and March must now be reported separately. **2) LHG** The degree of completeness for the reason for notification in the LHG data sources is very small (< 20%) across all years.

5.5.13 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	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 recoded 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 von IAB-BAMF-SOEP-ADIAB. This information does not refer to the start date of the episode, but to the start of the original time period.

5.5.14 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	The variable contains the person's status after leaving unemployment. Longer periods of illness can be identified via this variable. The values are classified as follows: • 1000s: measure (assisted employment) • 2000s: non-assisted employment • 3000s: training etc. • 4000s: self-employment • 5000s: exclusion • 6000s: other

5.5.15 Client profile (profil)

Variable label	client profile
Variable name	profil
Category	information on employment, benefit receipt and job search
Origin	ASU, MTH
Data type	numerical
Detailed description	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: • 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 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.
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.

5.5.16 Reason for end of previous employment (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, MTH
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.

5.5.17 Working hours of job application (arbzeit)

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

5.5.18 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, MTH
Data type	numerical
Detailed description	The variable has a different meaning depending on the data source. **1) LeH** The variable contains the residual entitlement to unemployment benefit that remains after the end of the current benefit receipt period. If the period of benefit receipt ends before the maximum duration of entitlement has been reached (e.g., due to taking up employment again), a residual entitlement remains which is equivalent to the duration of benefit entitlement that was not used up. If new entitlement is acquired within five years, the duration of the residual entitlement is added to the new duration of entitlement. However, the maximum duration of entitlement for the client's age is the upper limit. If no new entitlement is acquired, the residual entitlement can be used for benefits within four years on application. 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. **2) MTH** The variable contains the planned duration of the measure.
Notes on quality	**1) LeH** Over the course of a benefit receipt biography, a systematic development of the remaining entitlement is to be expected. This means that the remaining entitlement at the end of a benefit period minus the duration of the subsequent period should give the remaining entitlement after the subsequent period as long as no claims have arisen or expired in the meantime. However, there may be unexpected increases or decreases in the remaining entitlement. These are presumably mainly caused by corrections during the processing of benefit cases. Such corrections occur, for example, if a claim was not correctly determined at the start of the benefit case due to incomplete information or if a remaining claim that has not yet expired was not taken into account. However, these corrections are usually only administered for the current record, i.e. there is no correction of the complete case. In addition, in individual cases the information on entitlement durations may exceed the individual upper limit according to the statutory requirements.

5.5.19 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, MTH, XMTH
Data type	numerical
Detailed description	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 2.2.3. The type of institutions responsible for implementing SGB II may change over time (e.g., from a cooperation 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. ASU and XASU observations contain the type of institution that holds the records of the applicant pool data.

5.5.20 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, MTH, XMTH
Data type	numerical
Detailed description	The variable reports the start date of an uninterrupted sequence of periods of unemployment and is valid at the beginning of the observation. The following gaps do not result in an interruption of the period of unemployment: • any gap lasting seven days or less • periods of illness lasting up to 42 days

5.5.21 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, MTH, XMTH
Data type	numerical
Detailed description	The variable reports the duration (in days) of an uninterrupted sequence of periods of unemployment and is valid at the beginning of the observation. The following gaps do not result in an interruption of the period of unemployment: • any gap lasting seven days or less • periods of illness lasting up to 42 days 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". Prior to 1997, the value "0" does not mean that the individual was not unemployed, as the ASU/XASU sources are not available here.

5.6   Location data

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

Variable label	place of residence - district (Kreis)
Variable name	wo_kreis
Category	location data
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Hierarchy	federal state | district

Detailed description	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.
	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 and XASU sources, the place of residence applies to the whole period of the original observation. For the ASU, LeH, MTH and XMTH, 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.
	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.
Notes on quality	There are inaccuracies in the information provided for some employees with regard to where they live. The reporting requirement does not clarify which residence - main or secondary residence with predominant residence - is to be reported by the employer. In the employment history, this can lead to the impression of "long-distance commuters" between the registered main residence and the place of work, even though the person is in fact employed at the secondary residence, i.e. does not actually commute.
	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.
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.

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

Variable label	place of residence - federal state (Bundesland)
Variable name	wo_bula
Category	location data
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Hierarchy	federal state | district
Detailed description	This variable is an aggregation of the "district" variable to the 16 German federal states.
	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)

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

Variable label	place of residence - employment agency (Arbeitsagentur)
Variable name	wo_aa
Category	location data
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Hierarchy	regional directorate | employment agency
Detailed description	From 1999 onwards, this variable contains the agency district of the employment agency that is responsible for the employee's / BA client's place of residence. This information is determined from the residence address. For the LHG and XASU data sources, the place of residence is valid for the period of the original observation. In the case of the LeH, ASU, MTH and XMTH, 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.

	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).
Anonymisation	Owing to its particular sensitivity with regard to data privacy, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the area of the regional directorate in which the social security contributor's place of residence is located is shown.

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

Variable label	place of residence - regional directorate (Regionaldirektion)
Variable name	wo_rd
Category	location data
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
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.

5.7   Establishment variables

5.7.1   Classification of economic activities 73, groups (w73_3)

Variable label	classification of economic activities 73
Variable name	w73_3
Category	establishment variables
Origin	BHP
Data type	numerical
Hierarchy	division (1-digit code) | group (2-digit code) | class (3-digit code) of economic activity
Detailed description	This variable indicates the economic activity as a 3-digit code in accordance with the WS73 classification and is available from 1975 up to and including 2002.
	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 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.
	Each establishment is only assigned one code. The assignment to the relevant class of economic activity is carried out under consideration of the institutional orientation of the establishment.

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

Variable label	classification of economic activities 93, sub-classes
Variable name	w93_5

Category	establishment variables
Origin	BHP
Data type	numerical
Hierarchy	section (1-digit code) | division (2-digit code) | group (3-digit code) | class (4-digit code) | sub-class (5 digit code) of economic activity
Detailed description	This variable indicates the economic activity as a 5-digit code in accordance with the WZ93 classification and is available from 1999 up to and including 2003. WZ93 stands for the "Classification of Economic Activities for the Statistics of the Federal Employment Services, edition 1993" ("Klassifikation der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit, Ausgabe 1993"). The WZ93 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 1 („Nomenclature génerale des activités économiques dans les communautés européennes") which has four levels the first two of which are based on the international standard ISIC Rev. 3 („International Standard Industrial Classification of All Economic Activities"). Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.
Anonymisation	Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, the economic activity is only shown as the 3-digit code (w93_3).

5.7.3 Classification of economic activities 93, groups (w93_3)

Variable label	classification of economic activities 93, groups
Variable name	w93_3
Category	establishment variables
Origin	BHP
Data type	numerical
Hierarchy	section (1-digit code) | division (2-digit code) | group (3-digit code) | class (4-digit code) | sub-class (5 digit code) of economic activity
Detailed description	This variable indicates the economic activity as a 3-digit code in accordance with the WZ93 classification and is available from 1999 up to and including 2003. WZ93 stands for the "Classification of Economic Activities for the Statistics of the Federal Employment Services, edition 1993" ("Klassifikation der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit, Ausgabe 1993"). The WZ93 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 1 („Nomenclature génerale des activités économiques dans les communautés européennes") which has four levels the first two of which are based on the international standard ISIC Rev. 3 („International Standard Industrial Classification of All Economic Activities"). Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.

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

Variable label	classification of economic activities 03, sub-classes
Variable name	w03_5
Category	establishment variables
Origin	BHP
Data type	numerical
Hierarchy	section (1-digit code) | division (2-digit code) | group (3-digit code) | class (4-digit code) | sub-class (5-digit code) of economic activity
Detailed description	This variable indicates the economic activity as a 5-digit code in accordance with the WZ03 classification and is available from 2003 up to 2008. WZ03 stands for the "Classification of Economic Activities, Edition 2003" ("Klassifikation der Wirtschaftszweige Ausgabe 2003") of the Federal Statistical Office (eds.). Like the WZ93, the WZ03 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 1 (see description of variables w93_3, w93_5). The classifications of the economic activity have been updated, but the structure of the WZ93 has been largely retained. Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.

Anonymisation	Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, the economic activity is only shown as the 3-digit code (w03_3).

5.7.5 Classification of economic activities 03, groups (w03_3)

Variable label	classification of economic activities 03, groups
Variable name	w03_3
Category	establishment variables
Origin	BHP
Data type	numerical
Hierarchy	section (1-digit code)	division (2-digit code)	group (3-digit code)	class (4-digit code)	sub-class (5 digit code) of economic activity
Detailed description	This variable indicates the economic activity as a 3-digit code in accordance with the WZ03 classification and is available from 2003 up to 2008. WZ03 stands for the "Classification of Economic Activities, Edition 2003" ("Klassifikation der Wirtschaftszweige Ausgabe 2003") of the Federal Statistical Office (eds.). Like the WZ93, the WZ03 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 1 (see description of variables w93_3, w93_5). The classifications of the economic activity have been updated, but the structure of the WZ93 has been largely retained. Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.

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

Variable label	classification of economic activities 08, sub-classes
Variable name	w08_5
Category	establishment variables
Origin	BHP
Data type	numerical
Hierarchy	section (1-digit code)	division (2-digit code)	group (3-digit code)	class (4-digit code)	sub-class (5-digit code) of economic activity
Detailed description	This variable indicates the economic activity as a 5-digit code in accordance with the WZ08 classification and is available from 2008 onwards. WZ08 stands for the "Classification of Economic Activities, Edition 2008" ("Klassifikation der Wirtschaftszweige Ausgabe 2008") of the Federal Statistical Office (eds.). The WZ08 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 2. Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.
Anonymisation	Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise the economic activity is only shown as the 3-digit code (w08_3).

5.7.7 Classification of economic activities 08, groups (w08_3)

Variable label	classification of economic activities 08, groups
Variable name	w08_3
Category	establishment variables
Origin	BHP
Data type	numerical
Hierarchy	section (1-digit code)	division (2-digit code)	group (3-digit code)	class (4-digit code)	sub-class (5 digit code) of economic activity

Detailed description	This variable indicates the economic activity as a 3-digit code in accordance with the WZ08 classification and is available for the years since 2008. WZ08 stands for the "Classification of Economic Activities, Edition 2008" ("Klassifikation der Wirtschaftszweige Ausgabe 2008") of the Federal Statistical Office (eds.). The WZ08 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 2.
	Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.

5.7.8 Classification of economic activities 73, completed by extrapolation/imputation, groups (w73_3_gen)

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

5.7.9 Classification of economic activities 73, type of imputation w73_3, groups (group_w73_3)

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

5.7.10 Classification of economic activities 93, completed by extrapolation/imputation, groups (w93_3_gen)

Variable label	w93_3 completed by extrapolation/imputation
Variable name	w93_3_gen
Category	establishment variables
Origin	BHP
Data type	numerical
Hierarchy	section (1-digit code)	division (2-digit code)	group (3-digit code)	class (4-digit code)	sub-class (5 digit code) of economic activity

Detailed description	This variable indicates the economic activity as a 3-digit code in accordance with the WZ93 classification. From 1998 up to and including 2003, the variable contains the original values from w93_3. Before 1998 and after 2003, the information is either written back / continued or replaced with the help of recoding tables, so that the variable contains time-consistent information on the economic activity based on the economic activity classification WS93. A detailed description can be found in Eberle et al. (2011). Further information on the WS93 classification can be found in the description of variable w93_3.

5.7.11 Classification of economic activities 93, type of imputation, groups (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).

5.7.12 Classification of economic activities 08, completed by extrapolation/imputation, groups (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.

5.7.13 Classification of economic activities 08, type of imputation 08, groups (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).

5.7.14 Year of first appearance (grd_jahr)

Variable label	year of first appearance
Variable name	grd_jahr
Category	establishment variables
Origin	BHP
Data type	numerical
Detailed description	This variable indicates the first appearance of the establishment number in the dataset. If an establishment number in western Germany is only determined for the first time after 1975 (or after 1992 in eastern Germany), this variable could indicate the date when the respective establishment was founded. However, it could also be an establishment that has been in existence for a longer time but has been allocated a new establishment number following a change of owner or a change in the legal form of the establishment. (For the allocation of establishment numbers see Bundesagentur für Arbeit 2007, pp. 9-11). It is also possible that the establishment already existed before, but had no employees subject to social security, or from 1999 onwards, no marginal part-time workers. 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.

5.7.15 First appearance (grd_dat)

Variable label	first appearance
Variable name	grd_dat
Category	establishment variables
Origin	BHP
Data type	date
Detailed description	This variable indicates the first appearance of the establishment number in the BeH to the day. If an establishment number in western Germany is only determined for the first time after 1975, or after 1992 in eastern Germany, this variable could indicate the date when the respective establishment was founded. However, it could also be an establishment that has been in existence for a longer time but has been allocated a new establishment number following a change of ownership or a change in the legal form of the establishment. (For the allocation of establishment numbers see Bundesagentur für Arbeit 2007, pp. 9-11). It is also possible that the establishment already existed before, but had no employees subject to social security, or from 1999 onwards, no marginal part-time workers. 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.
Anonymisation	Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the year when the establishment number first appeared is shown (grd_jahr).

5.7.16 Year of last appearance (lzt_jahr)

Variable label	year of last appearance
Variable name	lzt_jahr
Category	establishment variables
Origin	BHP
Data type	numerical
Detailed description	This variable indicates the last appearance of the establishment number in the dataset (see Bender et. al. 1996). If the existence of an establishment number in the BHP already ends before 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). 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.

5.7.17 Last appearance (lzt_dat)

Variable label	last appearance
Variable name	lzt_dat
Category	establishment variables
Origin	BHP
Data type	date
Detailed description	This variable indicates the last appearance of the establishment number in the dataset to the day (see Bender et. al. 1996). If the existence of an establishment number in the BHP already ends before 2008, it could indicate the closure of the establishment. However, other possible rea-sons 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). 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.
Anonymisation	Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the year when the establishment number last appeared is shown (lzt_jahr).

5.7.18 Total number of employees (az_ges)

Variable label	no. employees
Variable name	az_ges
Category	generated establishment variables
Origin	BHP
Data type	numerical
Detailed description	This variable contains the total number of an establishment's employees reported to the social security agencies as of 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.

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

Variable label	no. full-time (regular workers + others)
Variable name	az_vz
Category	generated establishment variables
Origin	BHP
Data type	numerical
Detailed description	This variable contains the number of people in the establishment who are reported on 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.

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

Variable label	no. marginal part-time workers
Variable name	az_gf
Category	generated establishment variables
Origin	BHP
Data type	numerical
Detailed description	The number of employees in marginal part-time employment is generated using the person group code – values 109 and 209. This variable has only been contained in the dataset since 1999 as it has only been included in the social security notification procedure since that year.

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

Variable label	mean imp. wage all full-time employees
Variable name	te_imp_mw
Category	generated establishment variables
Origin	BHP
Data type	numerical
Detailed description	This variable contains the mean imputed gross daily wage of the full-time employees in an establishment. It does not include the wages of marginally part-time staff, apprentices or individuals participating in partial retirement schemes.
	The values are reported in euros for all years.
	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.
	In order to calculate the gross daily wage, the fixed period wage is divided by the number of calendar days in the period. To calculate the mean, these censored wages were imputed (see Section 3.1.3.3 in Schmucker et al. 2018). These data were then aggregated at establishment level. The values are rounded to two decimal places. However, due to the "storage type" in Stata, additional decimal places are displayed that are not correct.

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

Variable label	place of work - district (Kreis)
Variable name	ao_kreis
Category	location data
Origin	BHP
Data type	numerical
Hierarchy	federal state	district
Detailed description	The variable indicates the district (urban district or rural district) in which the employee's establishment is located. The first two digits of the 5-digit district code (Kreisschlüssel) show the code for the federal state (Bundesland, NUTS 1), positions 1-3 indicate the regional authority (Regierungsbezirk, NUTS 2). Federal states without a regional authority have a 0 in the third position. In the NUTS classification (Nomenclature des unités territoriales statistiques) of the European Union, districts correspond to the level NUTS 3.
	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 district code of the location of the establishment would change without the establishment concerned having relocated, if the territorial allocations were not updated.
Anonymisation	Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the federal state (ao_bula) is shown as regional information.

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

Variable label	place of work - federal state (Bundesland)
Variable name	ao_bula
Category	location data
Origin	BHP

Data type	numerical
Hierarchy	federal state | district
Detailed description	The variable indicates the federal state in which the establishment is located. This variable is generated from the district code (ao_kreis). The first two positions of the district code indicate the federal state (NUTS 1).

References

Bender, Stefan; Hilzendegen, Jürgen; Rohwer, Götz; Rudolph, Helmut (1996): Die IAB-Beschäftig-tenstichprobe 1975-1990. In: 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', 'aus-geübte Tätigkeit' sowie 'Schul- und Berufsabschluss' in der Beschäftigungsstatistik. Metho-denbericht, Bundesagentur für Arbeit – Statistik. URL: http://statistik.arbeitsagentur.de/Stati-scher-Content/Grundlagen/Methodenberichte/Beschaeftigungsstatistik/Generische-Publika-tionen/Methodenbericht-Neue-Erbebungsinhalte-Arbeitszeita-ausgeuebte-Taetigkeit-sowie-Schul-und-Berufsabschluss-in-der-Beschaeftigungsstatistik.pdf (Zugriff: 31.03.2016).

Brücker, Herbert; Rother, Nina; Schupp, Jürgen (2017): IAB-BAMF-SOEP Befragung von Geflüchte-ten 2016. Studiendesign, Feldergebnisse sowie Analysen zu schulischer wie beruflicher Quali-fikation, Sprachkenntnissen sowie kognitiven Potenzialen. IAB Forschungsbericht 13/2017.

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

Bundesagentur für Arbeit (Hrsg.) (2007): Handbuch für die Betriebsnummernvergabe und –pflege im Rahmen des Meldeverfahren zur Sozialversicherung. Ausgabe Dezember 2007, Nürnberg.

Bundesagentur für Arbeit (2009): Klassifikation der Wirtschaftszweige 1973, Nürnberg. URL: http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Wirt-schaftszweige/Klassifikation-der-Wirtschaftszweige-1973-2003/Klassifikationen-der-Wirt-schaftszweige-1973-2003-Nav.html (Zugriff: 21.04.2016)

Bundesagentur für Arbeit (2011): Klassifikation der Berufe 2010. Band 1: Systematischer und al-phabetischer Teil mit Erläuterungen, Nürnberg. URL: http://statistik.arbeitsagentur.de/Navi-gation/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/Naviga-tion/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB1975-1992/KldB1975-1992-Nav.html (Zugriff: 21.04.2016)

Christen, Peter (2012): Data Matching: Concepts and Techniques for Record Linkage, Entity Reso-lution, and Duplicate Detection. Berlin, Springer.

Cramer, Ulrich (1985): Probleme der Genauigkeit der Beschäftigtenstatistik. In: Allgemeines Statis-tisches 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.

Fellegi, Ivan; Sunter, Alan (1969): A Theory for Record Linkage. In: Journal of the American Statisti-cal Association, 64(328): S. 1183-1210.

Kosyakova, Yuliya, Olbrich, Lukas, Sakshaug, Joe W., & Schwanhäuser, Silvia (2019). Identification of interviewer falsification in the IAB-BAMF-SOEP Survey of Refugees in Germany (No. 201902_en). Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany].

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

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

Paulus, Wiebke; Matthes, Britta (2013): Klassifikation der Berufe * Struktur, Codierung und Umsteigeschlüssel. FDZ-Methodenreport, 08/2013 (de), Nürnberg.

Schnell, Rainer; Bachteler, Tobias; Reiher, Jörg (2005): MTB: Ein Record-Linkage-Programm für die empirische Sozialforschung. In: ZA-Information/Zentralarchiv für Empirische Sozialforschung, Nr. 56, S. 93-103.

Schmucker, Alexandra; Eberle, Johanna; Ganzer, Andreas; Stegmaier, Jens; Umkehrer, Matthias (2018): Betriebs-Historik-Panel 1975-2016. FDZ-Datenreport, 01/2018 (de), Nürnberg.

Statistisches Bundesamt (2002): Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ 93), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftklassifikationen/Content75/KlassifikationWZ93.html (Zugriff: 21.04.2016)

Statistisches Bundesamt (2003): Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftklassifikationen/Content75/KlassifikationWZ2003.html (Zugriff: 21.04.2016)

Statistisches Bundesamt (2008): Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftklassifikationen/Content75/KlassifikationWZ08.html (Zugriff: 21.04.2016)

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.

Appendix

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.

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
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 –übermittlungs-Verordnung –	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	Geschäftsbereich Daten- und IT-Management des Instituts für Arbeitsmarkt- und Berufsforschung	Data and IT Management of the IAB
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
FDZ	Forschungsdatenzentrum der Bundesagentur für Arbeit am IAB	Research Data Centre of the IAB
gAw	Träger mit getrennter Aufgabenwahrnehmung	Municipalities exercising their duties separately
gE	Gemeinsame Einrichtungen	Joint facility
IAB	Institut für Arbeitsmarkt- und Berufsforschung	Institute for Employment Research
IABS	IAB-Beschäftigtenstichprobe	IAB Employment Samples
IEB	Integrierte Erwerbsbiografien	Integrated Employment Biographies

ISIC	International Standard Industrial Classification of All Economic Activities	International Standard Industrial Classification of All Economic Activities
infas	Institut für angewandte Sozialwissenschaft, Bonn	Institute for Applied Social Sciences Bonn
LeH	Leistungsempfängerhistorik	Benefit Recipient History
LHG	Leistungs-Historik Grundsicherung	Unemployment Benefit II Recipient History
MTH	Maßnahmeteilnahmehistorik	Participants-in-Measures History File
NACE	Nomenclature générale des activités économiques dans les communautés eu-ropéennes	Nomenclature génerale des activités économiques dans les communautés euro-péennes
SGB	Sozialgesetzbuch	German Social Code
SIAB	Stichprobe der Integrierten Arbeitsmarktbiografien	Sample of Integrated Labour Market Biographies
VerBIS	Vermittlungs- und Beratungsinformationssystems	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
zkT	Zugelassene kommunale Träger	Authorised municipalities

Imprint

FDZ-Datenreport 15|2020 EN

Date of publication
26 March 2021

Download
http://doku.iab.de/fdz/reporte/2020/DR_15-20_EN.pdf

All publications in the series "FDZ-Datenreport" can be downloaded from
https://fdz.iab.de/de/FDZ_Publications/FDZ_Publication_Series/FDZ-Datenreporte.aspx

Website
https://fdz.iab.de/

Corresponding author
Dr. Sekou Keita
Phone: +499111791384
Email: sekou.keita@iab.de