FDZ-DATENREPORT

Documentation of labour market data

11|2019 EN NEPS-SC4 survey data linked to administrative data of the IAB (NEPS-SC4-ADIAB)

Nadine Bachbauer & Florian Zimmermann

NEPS-SC4 survey data linked to administrative data of the IAB (NEPS-SC4-ADIAB)

Nadine Bachbauer (IAB, LIfBi), Florian Zimmermann (IAB)

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

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

Contents

Zusammenfassung

Dieser Datenreport beschreibt die verknüpften Erhebungsdaten des Nationalen Bildungspanels der Startkohorte 4 „Schule und Ausbildung - Bildung von Schülerinnen und Schülern ab Klassenstufe 9" (NEPS-SC4) mit administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB).

Abstract

This data report describes the linked survey data of Starting Cohort 4 „School and Vocational Training: Educational Pathways of Students in Grade 9 and Higher" of the National Educational Panel Study (NEPS-SC4) with administrative data of the Institute for Employment Research (IAB).

Keywords

German administrative micro data, labour market data, data manual, National Educational Panel Study, survey data

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 data product "NEPS-SC4 survey data linked to administrative data of the IAB (NEPS-SC4-ADIAB 7517)", which is jointly conducted by the Leibniz Institute for Educational Trajectories (LIfBi) and the Institute for Employment Research (IAB). The linked data consist of information collected in interviews from participants of the so-called school cohort from grade 9 (SC4) of the National Education Panel (NEPS) and their personal data from the administrative data of the IAB, insofar as the respondents have given their consent to the linkage of the data and could be identified in the data of the IAB.

The NEPS data were collected from 2008 to 2013 as part of the Framework Programme for the Funding of Empirical Educational Research funded by the Federal Ministry of Education and Research (BMBF). Since 2014, the NEPS has been funded by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in close cooperation with a Germany-wide excellence network of partner institutes. Further background information can be found on the website of the NEPS.[1]

In terms of content, the Starting Cohort 4 of the National Educational Panel Study aims to trace the training paths taken by young people after completing the ninth grade of the general education school system. This includes paths to and through upper secondary level II as well as transitions into the vocational education system, higher education and the labour market.

The data enable the analysis of the competence equipment and development during the school and training career of adolescents and young adults. In addition, the data can be used to investigate the influences of school type transitions as well as transitions to further training, whether studies or vocational training. The analysis of (accumulated) experiences during vocational preparation measures or vocational training as well as the factors that constitute successful vocational training is also possible. In addition to these topics relating to vocational training, the NEPS also includes study conditions, the financial situation and prerequisites for studying, and the effects of these three components on the progress and success of studies. Finally, the transition into the labour market of young adults is also part of the main topics of NEPS-SC4.

The NEPS-SC4 survey includes a representative sample of students who attended the 9th grade of a general school in autumn 2010. The sample was drawn using multistage approaches. The first sample consists, in a first step, of a random selection of mainstream lower secondary schools. This is done by dividing the sample into five school types: Gymnasium, Real- und Hauptschule, Gesamtschule and schools with several courses of education. From these five school types, schools are drawn proportionally according to their number of classes, with an oversampling of schools whose graduates mainly enter the vocational training system. The second step involves the random selection of two ninth grades in the selected schools, provided that this grade strength is available. All students in the classes thus selected will be invited to participate in the study.

For the second sample special needs schools are selected at random according to size. A full survey of the 9th grades was then carried out at the special needs schools selected.

---

1 See https://www.neps-data.de/en-us/datacenter/dataanddocumentation/startingcohortgrade9.aspx

The target persons are interviewed and tested at the participating schools. In addition to the survey of students, the survey of contextual persons, parents, teachers and school principals is carried out.

Only the download version of the NEPS data is available for the linked data product. The remote or On-site version, where data is less anonymised, cannot be provided for NEPS-SC4-ADIAB.

The administrative data originate from the Integrated Employment Biographies (IEB) of the IAB. In the present case, they contain data on all persons interviewed under the NEPS Starting Cohort 4, who agreed to be linked and could be linked, if they have at least one of the following states 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 depicted exact to the day.

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 (Arbeitsuchendenhistoriken - ASU and XASU) are the data source for the periods of job search recorded by the BA or by municipal institutions responsible for implementing SGB II, while participation in employment and training measures is recorded in the Participants-in-Measures History Files (Maßnahmeteilnahmehistoriken – MTH and XMTH).

In addition to this personal data, information on the employing operators from the IAB's Establishment History Panel (BHP) (Schmucker et al., 2018) can also be provided.

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, a list of variables and the volume structure. A description of the individual data sources can be found in Section 2. Data preparation and sampling are discussed in Sections 3. Chapters 4 and 5 contain a description of data linkage and data quality, while the individual variables are described in Section 0.

1.2   Data use

1.2.1   Data access

The NEPS-SC4-ADIAB data are weakly anonymised and can only be evaluated during guest visits to one of the locations of the Research Data Centre of the Federal Employment Agency at the Institute for Employment Research (IAB-RDC) and subsequent remote data processing.

As a first step, an application has to be submitted to the IAB-RDC for data use. The IAB-RDC decides on the approval of the research project on behalf of and, if necessary, in consultation with the Federal Ministry of Labour and Social Affairs (BMAS). After approval, a data usage contract for the NEPS data must be concluded with the Research Data Centre of the LIfBi (LIfBi-RDC). Finally, as soon as the data use agreement for the NEPS data has been submitted to the IAB-RDC, a contract will be concluded between the institution of the researcher and the IAB. Details on how to apply for the data set and the data processing options can be found on the website of the IAB-FDZ. Access to the data thus only begins once the applicant has signed contracts with both research data centres.

To answer questions about this data product, users can contact the staff of the RDCs at IAB and LIfBi. The division of responsibilities between the two institutions includes that users with questions on NEPS data should contact the LIfBi-RDC exclusively, with questions on administrative data, technical aspects of data linkage or access to the data exclusively to the IAB-RDC. For general questions on NEPS, the NEPS Forum[2] can also be consulted in addition to the LIfBi-RDC.

1.2.2   Data management

The survey data of NEPS and the administrative data of the IAB are stored in separate files, which makes the structure of the data clear and saves storage capacity in the data management. The personal data from both data sources can be merged via the NEPS identification variable (ID_t), which is also included in the administrative personal data.

A further element of NEPS-SC4-ADIAB is the linkage file, which contains the personal ID as well as technical characteristics with information on whether the linkage with the administrative data was successful and which linkage method was used. The file also contains the respondents of NEPS Starting Cohort 4 for whom no administrative data exists. Thus, drop-out analyses can be carried out on the basis of all participants of NEPS-SC4.

The administrative data, which contain both German and English labels[3] analogous to the NEPS data, are structured in modules and stored in several files. One module, which is referred to as the "personal file", contains identifiers (system-free personal and company numbers), personal characteristics, information on employment, benefits, job search and participation in labour market measures, residence characteristics and technical characteristics. This personal data includes daily data for the period 1975 to 2017. The left margin of this data product is the year 2007, since the survey participants attended the 9th grade in 2010 and the birth cohorts range from 1991 to 1999. A second module, the basic establishment file, contains the establishment number, the year, information on the place of work and branches of the economy as well as further information on the establishment as at 30 June of the years 2007 to 2017. For eastern Germany, the establishment data are only available from 1992. Thus, administrative establishment and personal characteristics are stored in separate files (cf. Figure 1). In addition, the basic establishment file contains variables generated for sector classifications that are consistent over time (see Eberle et al., 2011). In addition, further establishment variables and extension modules of the Establishment History Panel (BHP) can be used, which are described by Schmucker et al. (2018). The linking of the administrative personnel and the establishment files are made via the establishment ID and the year

---

[2] See https://forum.neps-data.de/en-us/welcome.aspx
[3] The Stata commands label language en and label language de can be used to switch to English or German language labels.

of the data record. The variables marked with an "*" in the list of characteristics (see p. 19ff.) are contained in the basic establishment file.

Figure 1 Storage of the administrative data
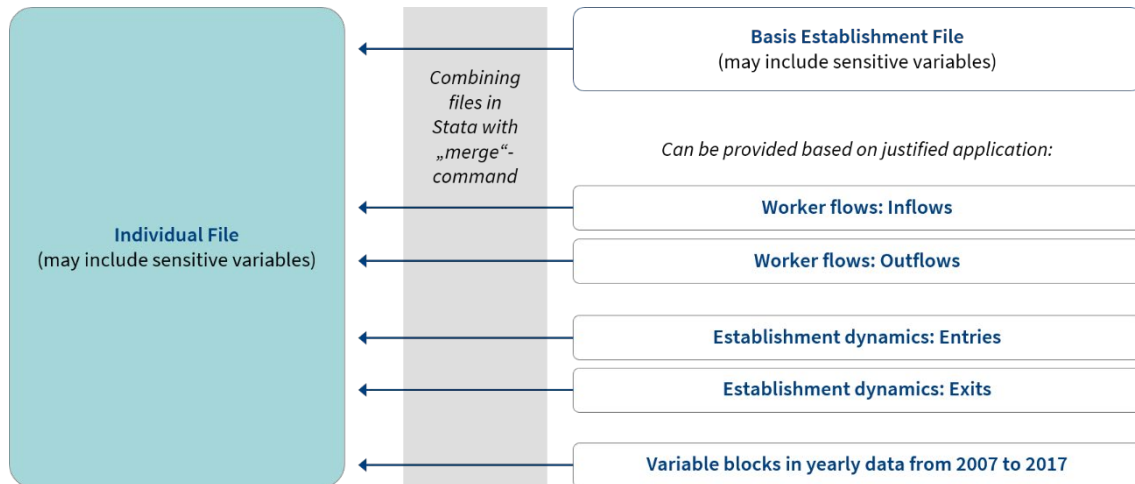
1.2.3 Sensitive variables and additional variables

Certain variables in the administrative data which make it possible to identify individuals or establishments are only disclosed in their original form if this is necessary for the analysis objective and is justified explicitly in the application for data access. The variables which are particularly sensitive from the viewpoint of data protection legislation are:

Administrative Individual File:

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

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.

On reasoned request, additional establishment variables can be provided as blocks of variables from the Establishment History Panel (BHP) (see Figure 1). Further information on the possible additional blocks of the BHP that can be requested can be found under https://fdz.iab.de/en/FDZ_Overview_of_Data/working_tools.aspx.

The NEPS data are also available in different degrees of anonymity. The data available for NEPS-SC4-ADIAB is the download version with the highest degree of anonymity. This version contains several variables, each marked with the suffixes "_R" or "_O" in their names, which have been anonymised by aggregation or other measures. The full information on these variables is only available in the RemoteNEPS or the On-site version. For privacy reasons, neither the RemoteNEPS nor the On-site version can be made available for NEPS-SC4-ADIAB, not even with a separate application. There is thus no possibility to use variables from the NEPS survey that are not included in the download version. This applies primarily to regional data such as administrative regions and districts as well as country identifiers such as native language or country of origin.

1.2.4   File names of the administrative data and the Linkage File

Individual File

      NEPS-SC4-ADIAB_7517_v1.dta

Basis Establishment File

      NEPS-SC4-ADIAB_7517_v1_bhp_basis_v1.dta

Linkage File

      NEPS-SC4-ADIAB_7517_v1_schluessel.dta

Extension Files BHP

Core datasets with blocks of variables

      NEPS-SC4-ADIAB_7517_v1_bhp_v1_jjjj.dta, jjjj = 2007 - 2017

Worker flows

      NEPS-SC4-ADIAB_7517_v1_bhp_inflow_v1.dta

      NEPS-SC4-ADIAB_7517_v1_bhp_outflow_v1.dta

Establishment dynamics

      NEPS-SC4-ADIAB_7517_v1_bhp_entry_v1.dta

      NEPS-SC4-ADIAB_7517_v1_bhp_exit_v1.dta

1.2.5   File names of the NEPS Starting Cohort 4

Table 1 data structure of NEPS-SC4

Dataset	File name	Content and special features
Profile data	CohortProfile.dta	Information on participation, interview date, availability of survey and competence data; initial file for merging information (identifiers); longitudinal data in long format
Methods data of the survey	TargetMethods.dta	Information on the field characteristics and data collection of the target person (response rate, contact, duration, interviewer, incentives); longitudinal data in long format
	ParentMethods.dta	Information on the field characteristics and data collection of the parents (response rate, contact, duration, interviewer, incentives); longitudinal data in long format
Weighting data	Weights.dta	Information on sample characteristics and stratification including weighting variables; longitudinal data in long format
	RepWeights.dta	Replication weights for the target person and the institutions
Survey data	pTarget.dta	Information from the interviews with the target person (students, PAPI); longitudinal data in long format
	pTargetCATI.dta	Information from the interviews with the target person (apprentices, employed persons, university students, students outside the general education system, CATI); longitudinal data in long format
	pTargetCAWI.dta	Information from the interviews with the target person (CAWI); longitudinal data in long format
	pParent.dta	Information from the interviews with the parents; longitudinal data in long format
Context data	pCourseClass.dta	Information from the (class) teacher survey on the class context; longitudinal data in long format
	pCourseGerman.dta	Information from the survey of subject teachers (German); longitudinal data in long format
	pCourseMath.dta	Information from the survey of subject teachers (Maths); longitudinal data in long format
	pEducator.dta	Information from the (class) teacher survey on teachers; longitudinal data in long format
Competence data	xTarget Competencies.dta	Information on the competence tests carried out in waves 1-3, 5, 7 and 10; longitudinal data in wide format
	xTargetSpecial-NeedsCompeten-cies.dta	Information on the competence tests with special needs students carried out in waves 1-3, 5, 7 and 10; longitudinal data in wide format
Life-course spell data	Biography.dta	Integrated and smoothed life-course data from different areas of life; summary of all complete, harmonised and right-censored episode spells

Education data	Education.dta	Information on the transitions in the target person's (educational) career (e.g. CASMIN, ISCED-97); dataset generated from the spell records for education and training
School education spell data	spSchool.dta	Information on the general education history, from school enrolment to the (anticipated) time of completion
	spParentSchool.dta	Information on the general education history of the parents, from school enrolment to the (anticipated) time of completion
Vocational preparation spell data	spVocPrep.dta	Information on episodes of vocational preparation after general education
Further vocational training spell data	spVocTrain.dta	Information on all further training that the target person has ever completed (vocational training, studies)
Employment spell data	spEmp.dta	Information on regular work episodes including internships; beginning of a new episode in the event of a change of employer, employment or career break
Unemployment spell data	spUnemp.dta	Information on episodes of unemployment irrespective of registration as unemployed
Military or civilian service spell data	spMilitary.dta	Information on episodes of military and civilian service as well as periods used for voluntary work in the social or environmental sectors
Children spell data	spChild.dta	Information about the target person's biological, foster and adoptive children as well as all children who live or have lived with the target person
Parental leave spell data	spParLeave.dta	Information on each child from spChild (excluding deceased children) and information on whether or not the target person has taken parental leave
Child cohabitation spell data	spChildCohab.dta	Information on the target person's history of cohabitation with children
Sibling spell data	spSibling.dta	Information on the care and education of the siblings from the parent survey
Gap spell data	spGap.dta	Information on gaps in the recording of the individual life course; identified by test module
	spParentGap.dta	Information on gaps in the recording of the parental life course; identified by test module
Occupational course spell data	spCourses.dta	Information on courses and training courses attended during the last 12 months during an employment, unemployment, parental leave, military/civilian service or gap episode
Foreign school qualification spell data	spSchoolExt Exam.dta	Information about school leaving certificates acquired abroad
Foreign vocational qualification spell data	spVocExtExam.dta	Information on certificates of completion (of training) obtained abroad
Voluntary work spell data	spVolunteer- Work.dta	Information on voluntary activities (e.g. duration, requirements, content of the activity)
Edition data	EditionBackups.dta	Information about editions made in the data sets

The merging matrix on the NEPS website indicates which variables in the respective datasets act as unique identifiers for merging information. Further information on the individual datasets can be found in the Data Manual of SC6, which is also available on the website.

1.3 Profile

Topics/variable groups	Survey Data Longitudinal data on sociodemography; education and employment biography; competence development, education processes, education decisions and returns to education in formal, non-formal and informal contexts Administrative Personal Data **Employee History (BeH):** Yearly reports and deregistration from social insurance for employees who are subject to social security contributions or are marginally employed **Benefit Recipient History (LeH):** Information on benefit receipt according to SGB III for recipients of unemployment benefits, unemployment assistance and maintenance allowance **Unemployment Benefit II Recipient History (LHG):** Data on individuals who receive benefits to ensure their subsistence according to SGB II (types of agencies: ARGE/joint institutions, separate institutions/separate responsibility, authorised communal agencies) **Jobseeker History (ASU):** Information on job search **Jobseeker History from XSozial-BA-SGB II (XASU):** Information on job search which is transmitted by authorised communal agencies to the BA via the transmission standard XSozial-BA-SGB II **Participants-in-Measures History Files (MTH):** Information on participation in employment and training measures (not including measures of authorised municipalities) **Participants-in-Measures History File from XSozial-BA-SGB II (Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II - XMTH):** Information on (planned) participation in measures of active labour-market policy (not measures run by the authorised communal agencies in the sphere of the SGB II) Administrative Establishment Data **Core dataset:** Establishment variables (e.g. artificial 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 employees, structure of the employees by education and vocational training, structure of the employees by classification of occupation (Blossfeld classification), average wage of full-time employees **Extension 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 **Extension 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	Survey Data: People who attended the 9th grade of a general school in Germany in autumn 2010. Oversampling of schools whose graduates mostly enter the vocational training system and separate sample of special needs schools Administrative Personal Data: Employees subject to social security contributions (also marginally employed from 1999 onwards), benefit recipients, job seekers, participants in measures
Case numbers	Survey Data: Wave 1: 16.106 participants (linked: 0[4] persons) Wave 2: 15.215 participants (additionally linked: 0) Wave 3: 14.011 participants (additionally linked: 1.270) Wave 4: 1.351 participants (additionally linked: 0) Wave 5: 12.982 participants (additionally linked: 5.012) Wave 6: 5.392 participants (additionally linked: 0) Wave 7: 11.830 participants (additionally linked: 200) Wave 8: 9.871 participants (additionally linked: 2.318) Wave 9: 9.044 participants (additionally linked: 369) Wave 10: 7.986 participants (additionally linked: 0) Administrative Data: 9.169 persons 104.902 original records 167.176 records without any overlaps (after episode splitting)
Period	Survey Data: Wave 1: Autumn 2010 Wave 2: Summer 2011 Wave 3: 2011/2012 Wave 4: Spring 2012 Wave 5: 2012/2013 Wave 6: Spring 2013 Wave 7: 2013/2014 Wave 8: 2014/2015 Wave 9: 2015/2016 Wave 10: 2016/2017 Administrative Data: The period depends on the data source. BeH: 01.01.2007 - 31.12.2017 (2014: 36-month file 2015 and 2016: 18-month file, 2017: 6-month file) LeH: 01.01.2012 - 31.12.2017 ASU: 01.01.2007 - 31.12.2017 LHG: 01.01.2007 - 31.12.2017

---

[4] In waves 1 and 2, none of the target persons interviewed was of legal age, so the question of consent to link BA data was not asked.

	XASU: 01.01.2008 - 31.12.2017
	MTH: 01.01.2007 - 31.12.2017
	XMTH: 01.01.2009 - 31.12.2017
Time reference	Survey Data: Time of survey; spell and event data retrospectively for the first survey, subsequently since the last survey Administrative Personal Data: Daily employment biographies Administrative Establishment Data: Reporting date 30 June
Regional structure	Federal states, districts (Kreise)
Territorial status	Adjusted to the territorial status as of 31.12.2017
Survey design	Survey Data:
	Wave 1: Survey of students, teachers and school principals (PAPI), competence testing of students (PAPI; mathematics/ICT literacy/natural sciences/hearing comprehension/reading speed/metacognition) and parent interview (CATI)
	Wave 2: Survey of students (PAPI) and competence testing of students (PAPI; Reading/Metacognition/Cognitive Basic Skills/Hearing comprehension)
	Wave 3: Survey of students, teachers and school principals (PAPI), competence testing of students (PAPI; English reading competence/metacognition), telephone interview (CATI) or personal interview (CAPI) of school leavers
	Wave 4: Telephone interview (CATI) of school leavers
	Wave 5: Survey of students and school principals (PAPI), competence testing of students (PAPI; natural sciences/metacognition), parent interview (CATI), telephone interview (CATI) or personal interview (CAPI) of school leavers as well as additional online surveys of school leavers and school changers, if applicable
	Wave 6: Telephone interview (CATI) of school leavers
	Wave 7: Survey of students and school principals (PAPI), competence testing of students and partly of the school leavers (PAPI; mathematics/reading/ICT literacy/English/scientific-critical thinking/metacognition), parent interview (CATI), telephone interview (CATI) or personal interview (CAPI) of school leavers as well as additional online surveys of school leavers and school changers, if applicable
	Wave 8: Repeated survey pre-wave (PAPI, CATI or CAPI) as well as additional online survey of school leavers and school changers, if applicable
	Wave 9: Telephone interview (CATI) or face-to-face interview (CAPI) as well as additional online interview (CAWI)
	Wave 10: Face-to-face interview (CAPI) with competence testing (mathematics/reading/cognitive basic skills); task processing using technology-based testing on a laptop (TBT); alternatively telephone interview (CATI) as well as additional online survey (CAWI) of students, trainees in the last year of vocational training and students at general schools
	Administrative Data: Identification of the respondents in the Integrated Employment Biographies of the IAB

Institutions involved	Survey Data: Contracting body: Leibniz-Institute for Educational Trajectories e.V. (LIfBi) at the Otto-Friedrich-University of Bamberg Implementation of the survey: infas Institute for Applied Social Sciences Administrative Data: Social insurance agencies, Federal Employment Agency
Frequency of data collection	Survey Data Approx. annual Administrative Data Ongoing
File format and size	Survey Data Stata; 33 single data files between 87 KB und 163,2 MB Administrative Data Stata; Personal data set: 12,3 MB; Establishment data set: 0,31 MB
File organisation	Survey Data The data is stored in several files containing personal data stored as cross-sectional or longitudinal data (wave- or episode-specific). Administrative Data The data is stored in two files. The first includes personal and the other establishment information. Upon reasoned request, further files with additional establishment information are provided. Linkage Data Furthermore there is a file containing the identifier variable as well as information on the matching success.
Data access	On-site at IAB-RDC, remote data access after previous On-site visit
Degree of anonymisation	Weakly anonymised
Sensitive variables	Month of birth (gebmon), Nationality (nation), Occupational sub group – current/most recent (beruf2010_4), Place of residence: district (Kreis) (wo_kreis), Place of residence: employment agency (wo_aa), Place of work: district (Kreis) (ao_kreis), Sub-class of economic activity (w93_5), Sub-class of economic activity (w03_5), Sub-class of economic activity (w08_5), Date of first appearance (grd_dat), Date of last appearance (lzt_dat)
Citation of the data and data documentation	Data: "The data basis of this report is the NEPS-SC4 survey data linked to administrative data of the IAB (NEPS-SC4-ADIAB). The data were accessed via a research visit at the Research Data Centre (RDC) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB-RDC) and/or via controlled remote data processing at the RDC. " DOI: 10.5164/IAB.NEPS-SC4-ADIAB7517.de.en.v1 Datendokumentation: Bachbauer, Nadine; Zimmermann, Florian (2019): NEPS-SC4 survey data linked to administrative data of the IAB (NEPS-SC4-ADIAB 7517). FDZ-Datenreport, 11/2019 (de), Nürnberg. DOI: 10.5164/IAB.FDZD.1911.en.v1
Dataset version	NEPS-SC4 survey data linked to administrative data of the IAB (NEPS-SC4-ADIAB 7517); DOI: 10.5164/IAB.NEPS-SC4-ADIAB7517.de.en.v1

1.4 List of variables

The overview of variables in Table 3 lists the variable names and the longer descriptions of variables of the administrative data. Regarding the establishment data, only the variables of the Basis Establishment File are listed, as in chapter 6. The other establishment variables as well as the extension modules are described in detail in Schmucker et al. (2016)

It also provides an overview of whether and how well variables are filled in the individual data sources. Table 2 illustrates the meaning of the shading, which indicates the degree of completeness per variable and source in Table 3.

Table 2 Degrees of completeness of variables

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

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

Table 3 List of variables of the administrative data with degree of completeness

List of variables	Module	Page	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Identifiers		**42**							
Individual ID (ID_t)		42	h	h	h	h	h	h	h
Establishment ID (betnr)	*	42	h	l	l	l	l	l	l
Generated technical variables		**43**							
Counter per person (spell)		43	h	h	h	h	h	h	h
Source of spell (quelle)		44	h	h	h	h	h	h	h
Year (jahr)	*	44	l	l	l	l	l	l	l
Period of validity		**44**							
Original start date (begorig)		44	h	h	h	h	h	h	h
Original end date (endorig)		45	h	h	h	h	h	h	h

List of variables	Module	Page	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Episode start date (begepi)		45	h	h	h	h	h	h	h
Episode end date (endepi)		45	h	h	h	h	h	h	h
Personal information		**46**							
Gender (frau)		46	h	h	h	h	h	h	h
Year of birth (gebjahr)		46	h	h	h	h	h	h	h
Month of birth (gebmon)		46	h	h	h	h	h	h	h
Nationality (nation)		47	h	h	h	h	h	h	h
Nationality, grouped (nation_gr)		47	h	h	h	h	h	h	h
Marital status (famst)		47	l	h	h	h	h	m	h
Number of children (kind)		48	l	h	h	m	m	m	m
Vocational training (ausbildung)		48	m	l	l	h	m	h	m
Vocational training (imputed) (ausbild-ung_imp)		51	h	l	l	l	l	l	l
School leaving qualification (schule)		51	m	l	l	h	m	h	m
Information on employment, benefit receipt and job search		**52**							
Daily wage/daily benefit (tentgelt)		52	h	h	l	l	l	l	l
Occupation - current/most recent (KldB 1988) (beruf)		53	h	l	l	m	l	m	l
Occupational group - current/most recent (KldB 2010), 3-digit (beruf2010_3)		54	h	m	l	m	m	m	m
Occupational sub-group - current/most recent (KldB 2010), 4-digit (beruf2010_4)		54	h	m	l	m	m	m	m
Level of requirement - current/most recent job (KldB 2010) (niveau)		55	h	m	l	m	m	m	m
Part-time (teilzeit)		56	h	l	l	l	l	l	l
Occupational status and working hours (stib)		56	h	l	l	l	l	l	l
Employment status (erwstat)		57	h	h	h	h	h	h	h
Transition zone (gleitz)		59	h	l	l	l	l	l	l
Temporary agency work (leih)		59	h	l	l	l	l	l	l
Fixed-term contract (befrist)		59	h	l	l	l	l	l	l
Reason of cancellation/notification/ter-mination (grund)		60	h	h	m	h	m	l	l
Employment status prior to job search (estatvor)		61	l	l	l	h	m	l	l
Employment status after job search (es-tatnach)		61	l	l	l	m	l	l	l
Client profile (profil)		62	l	l	l	m	l	m	l
Reason for end of previous employment (art_kuend)		62	l	l	l	m	l	m	l
Working hours of job application (arbzeit)		63	l	l	l	m	l	m	l
Residual claim/planned duration (restanspruch)		63	l	h	l	l	l	h	l
Type of provider (traeger)		64	l	l	h	h	h	h	h

List of variables	Module	Page	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Start date of unemployment (alo_beg)		64	m	m	m	m	m	m	m
Duration of unemployment (alo_dau)		64	h	h	h	h	h	h	h
Location data		**65**							
Place of residence - district (Kreis) (wo_kreis)		65	h	h	h	h	h	h	h
Place of residence - federal state (Bundesland) (wo_bula)		66	h	h	h	h	h	h	h
Place of residence - employment agency (Arbeitsagentur) (wo_aa)		66	h	h	h	h	h	h	h
Place of residence - regional directorate (Regionaldirektion) (wo_rd)		67	h	h	h	h	h	h	h
Establishment variables		**67**							
Classification of economic activities 73 (w73_3)	*	67	h	l	l	l	l	l	l
Classification of economic activities 93, sub-classes (w93_5)	*	67	h	l	l	l	l	l	l
Classification of economic activities 93, groups (w93_3)	*	68	h	l	l	l	l	l	l
Classification of economic activities 03, sub-classes (w03_5)	*	68	h	l	l	l	l	l	l
Classification of economic activities 03, groups (w03_3)	*	69	h	l	l	l	l	l	l
Classification of economic activities 08, sub-classes (w08_5)	*	69	h	l	l	l	l	l	l
Classification of economic activities 08, groups (w08_3)	*	70	h	l	l	l	l	l	l
w73_3 completed by extrapolation/imputation (w73_3_gen)	*	70	h	l	l	l	l	l	l
Type of imputation w73_3 (group_w73_3)	*	70	h	l	l	l	l	l	l
w93_3 completed by extrapolation/imputation (w93_3_gen)	*	71	h	l	l	l	l	l	l
Type of imputation w93_3 (group_w93_3)	*	71	h	l	l	l	l	l	l
w08_3 completed by extrapolation/imputation (w08_3_gen)	*	71	h	l	l	l	l	l	l
Type of imputation w08_3 (group_w08_3)	*	72	h	l	l	l	l	l	l
Year of first appearance (grd_jahr)	*	72	h	l	l	l	l	l	l
First appearance (grd_dat)	*	72	h	l	l	l	l	l	l
Year of last appearance (lzt_jahr)	*	73	h	l	l	l	l	l	l
Last appearance (lzt_dat)	*	73	h	l	l	l	l	l	l
Total number of employees (az_ges)	*	74	h	l	l	l	l	l	l
Number of full-time employees (regular workers + others) (az_vz)	*	74	h	l	l	l	l	l	l
Number of employees in marginal part-time employment (az_gf)	*	74	h	l	l	l	l	l	l

List of variables	Module	Page	BeH	LeH	LHG	ASU	XASU	MTH	XMTH
Mean imputed wage all full-time employees (te_imp_mw)	*	74	h	l	l	l	l	l	l
Place of work - district (Kreis) (ao_kreis)	*	75	h	l	l	l	l	l	l
Place of work - federal state (Bundesland) (ao_bula)	*	75	h	l	l	l	l	l	l

1.5 Volume structure

Table 4 Volume structure

No. of cases	before splitting	after splitting
BeH	63,751	84,946
LeH	3,276	6,014
LHG	12,412	25,765
ASU	16,374	32,808
XASU	3,361	6,592
MTH	5,329	10,009
XMTH	399	1,042
Total number of records	104,902	167,176
Individuals		9,169

2 Data sources

2.1 NEPS-SC4 survey data (NEPS-SC4)

The NEPS-SC4 sub-study "School and education - education of students from grade 9 onwards" provides a data basis for research on educational and training pathways, entry into employment and competence development in the life course of adolescents and young adults. The aim of this sub-study is to examine the various paths taken by young people in the general education school system after completing grade 9. In addition to the transition to secondary level II, vocational training or a course of study, the NEPS study also focuses on the entry into the labour market. The existing gaps in knowledge about the determinants of competence acquisition and the development of various competences depending on the type of school can also be closed with this partial study. To this end, competence measurements will be carried out in the areas of "language", "mathematics", "natural sciences", "ICT" and "cognitive problem-solving skills".

In terms of content, the data collected should make it possible, among other things,

- to trace the educational pathways of adolescents and young adults throughout their (educational) history and beyond their entry into working life;

- to identify the determinants of the acquisition of competences and their different development, taking into account the type of school;

- to describe the frequency and influencing factors of school form changes;

- to evaluate the success or failure of the preparation for university studies within the framework of the upper secondary school;

- to explain the factors influencing the access to the vocational education system and the impressions and difficulties experienced by adolescents during the transition from the general school system to the vocational education system;

- to analyse the experience gained in vocational preparation measures or in vocational training and to determine the key success factors of vocational education;

- to assess the reasons for taking up a course of university studies;

- to generate empirical results on the effects of study conditions, the financial situation and the study prerequisites on the course of studies and academic success;

- and to shed light on the mechanisms and variables influencing the transition of young adults into the labour market.

2.2 Administrative Data of the Integrated Employment Biographies (IEB)

The administrative individual data were drawn from the Integrated Employment Biographies (IEB) of the IAB. They unite data from different data sources, each of which may contain information from different administrative procedures. In addition, some supplementary variables from these data sources, which are not part of the IEB, are incorporated into the administrative individual data. Figure 2 illustrates the data flows that lead to the NEPS-SC4-ADIAB and their relationship to other FDZ data products.

Figure 2 Data sources of the NEPS-SC4-ADIAB 7517
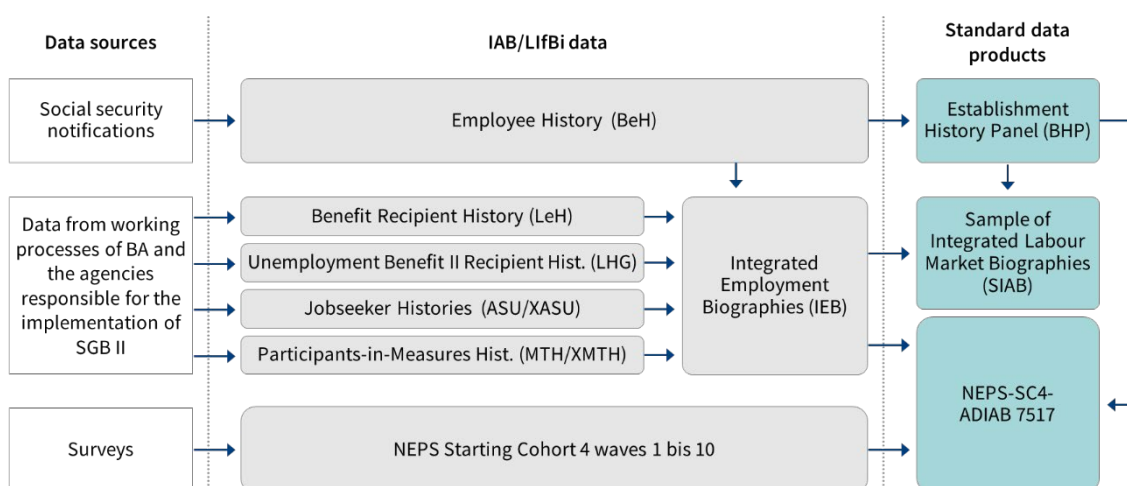
2.3 Employee History (BeH)

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

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

When linking individual data with establishment data, one has to take into account that the variables in both the Basis Establishment File and the BHP Extension Files are aggregated on 30 June of a year. Unlike the data on individuals, the establishment variables are therefore not spell data but are only valid on 30 June precisely each year.[6]

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,

---

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

[6] 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.

the two files can be linked using the "merge" command in connection with the relevant paths (see Box 1).

Box 1 Example code for merging individual and establishment data in Stata 14

```
use NEPS-SC4-ADIAB_7517_v1.dta, clear
gen int jahr = year(begepi)
sort betnr jahr
merge m:1 betnr jahr using NEPS-SC4-ADIAB_7517_v1_bhp_basis_v1.dta
```

2.4   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.

2.5   Unemployment Benefit II Recipient History (LHG)

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

- 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.[8] Authorised municipalities use various IT procedures of their own and transmit their data to the BA by means of the XSozial-BA-SGB II standard once a month. Both of the procedures are used by municipalities with separated responsibilities. The different data standards affect the scope and quality of the data supplied.

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

2.6 Jobseeker Histories (ASU/XASU)

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

2.7 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.

---

[7] 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.

[8] 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.

3   Data preparation and sampling procedure

3.1   Corrections and validation procedures

3.1.1   NEPS-SC4 survey data

Detailed information on sampling, data collection, the instruments used, and the structure and processing of NEPS-SC4 survey data can be found on the NEPS website. In addition to a series of reports and tools for handling the data, the so-called "semantic data structure file" is also available, which provides a first insight into the datasets of the SC4 including variable names, variables and value labels (but no observations)[9].

3.1.2   Administrative data from the Integrated Employment Biographies

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

3.1.3   Employee History (BeH)

- To capture a person group that is as constant as possible over time, some person groups for which data are not available throughout the entire observation period are excluded. From the reporting year 2011 onwards, the BeH data originate from newly designed source data. As a result, a number of person groups have been introduced or reactivated as they are classified by the BA statistics as being subject to social security contributions. The person groups 101 - 107, 111 - 114, 118, 119, 120, 140, 141, 142, 143, 149, 201 and 203 - 205 are therefore contained from that time onwards as well as the two groups 109 and 209, which indicate people in marginal part-time employment. Groups that are not included are, for example, people in short-term employment, i.e. person groups 110, 202 and 210.

- Person groups 123, 124 and 127 have been newly introduced in 2011.

- For data protection reasons, the person groups 107, 111, 113, 114, 127 and 204 are combined to form the person group "other workers" (599).

---

[9] See https://www.neps-data.de/en-us/home.aspx.

- From the reporting year 2012 onwards apprentices were included as the new person groups 121 and 122.

- Observations with no details on earnings, and the value 101 for the person group variable, and the value 50 for the reason for notification (annual notification) are not incorporated into the IEB.

- Gender and date of birth are taken from the Data Warehouse (DWH) of the BA. This information is harmonised across data sources.

- The territorial allocations for place of work and place of residence are updated to the status as of 31 December 2017.

3.1.4   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.

3.1.5   Unemployment Benefit II Recipient History (LHG)

- Observations without a BA client number are deleted.

- Cancelled data records are not used.

- It only contains observations of people who are capable of work and people under the age of 65.

- In each case, non-overlapping periods of benefit entitlement of a person in a certain benefit unit (BG) are created. New observations are begun for the following administrative reasons:
  - on certain birthdays of members of the BG that are stipulated by law and relevant for structural changes in the benefit unit (14, 15, 18 and 25) and the individual retirement age of members of the BG,
  - when the structure of the benefit unit changes (e.g. due to entries/exits),
  - when there are changes in a variable of the BG client and
  - at the beginning and the end of a case of benefit sanctions for observations from 1 April 2006 onwards. It must be taken into account, however, that it is not possible to identify the duration or type of sanction or the time when it was imposed or when it began on the basis of the data. The reason for this is that there is no corresponding variable or value that indicates the start, type or duration of the sanction.

- For the reason mentioned above, all individual-related variables that are available for the LHG source are valid for the entire duration of the observation.

- Double notifications due to the territorial reforms in 2009/2011 and the reorganisation of the institutions in 2012 (see Section 5.7.1) were corrected as far as possible.

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

3.1.6 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.

3.1.7 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.

3.1.8 Participants-In-Measures History File from 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.

3.1.9   SGB II anonymisation

In order to reduce the risk of de-anonymisation, only the year of birth is available in the SIAB 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 and/or the day before, or start again on the birthday.

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

3.2   Episode splitting

The administrative individual data are available with "split" episodes. If observations overlap within an account, these observations are replaced by artificial observations with new dates so that completely parallel periods and non-overlapping periods are created. This increases the number of observations (see Figure 3).

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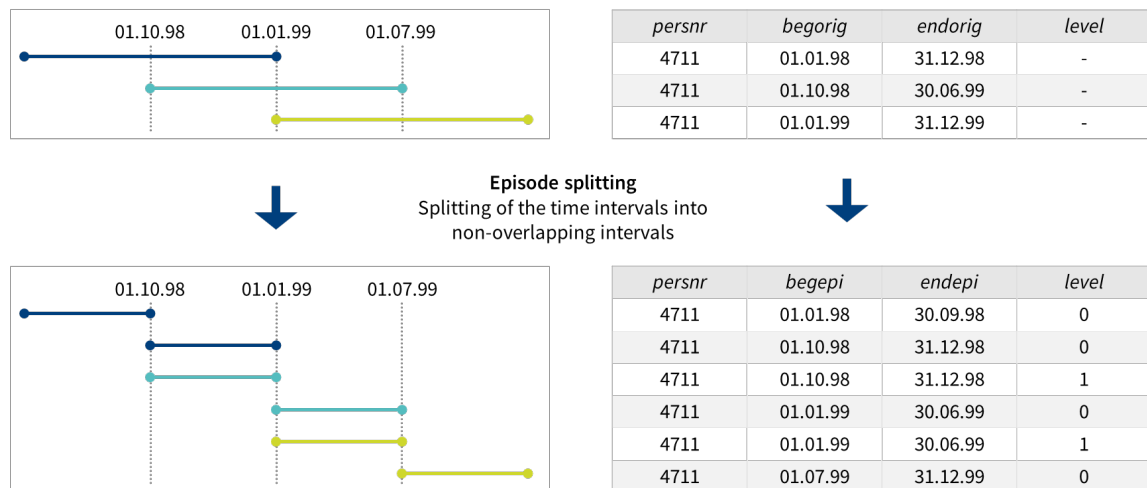
The original date variables for the beginning and the end of the original observation (begorig and endorig) are retained, the variables 'start date of the split episode' and 'end date of the split epi- sode' (begepi and endepi) mark the beginning and the end of the split episodes. It is possible to establish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

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

It is advisable to sort entirely parallel observations generated by the splitting procedure in a con- sistent manner. The variables 'observation counter per episode' (level2) and 'observation counter per episode and source' (level1) that were previously contained in the SIAB can be generated using the following Stata commands if required (see Box 2).

Box 2 Example code to create additional observation counters in Stata 14

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

3.3 Missing values

In the data missing values are coded as follows:

Term	Value	Description
No (valid) details available	.z	Values of a variable that are not systematically missing, i.e., the variable is available in principle for the data source, but no de- tails are available for the value considered or cannot be inter- preted reasonably.
Systematically not available	.n	A variable is not available in principle for a data source (dark grey cells in the overview of variables in Section 1.4) or is not available for a certain period.

4 Data Linkage

4.1 Respondents' informed consent for the data linkage

The starting point for linking the survey and process data are the respondents of the NEPS-SC4. During the interview, they were asked for their consent to their survey data being linked with the administrative data available about them at the IAB for research purposes. The questionnaire section with the declaration of consent reads as follows:

"In order to keep the following interview sections as short as possible, we would like to incorporate data held by the "Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit" (Institute for Employment Research of the Federal Employment Agency) in Nuremberg into the analysis of this questionnaire. This includes, for instance, information relating to previous employment relationships, phases of unemployment, participation in training measures during unemployment as well as company characteristics. I would ask you to give your official consent for the transfer and incorporation of this data into the interview data. If this information is evaluated, we guarantee that it is done so in strict compliance with all data protection regulations and that no information is passed on to third parties. Your consent is, of course, voluntary. You can withdraw this consent at any time by informing the contact person at infas. Do you consent to the transfer of data held by the "Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit" (Institute for Employment Research of the Federal Employment Agency)?"

Respondents had to agree to the above question in order to become part of the sample for the linkage. The question concerning consent to the linkage was asked in waves 3 through 10, with people being asked this question again only if they had previously not agreed to the linkage.

4.2 Origin of the linkage variables

The following personal variables from both data sources were used to link the survey data and the administrative IAB data of the individuals interviewed in NEPS-SC4 who had given their consent:

- First and last name
- Date of birth
- Gender
- Address (postal code, town, street, house number)

On the part of the NEPS survey, these variables were transmitted to the IAB by the Institute for Applied Social Sciences (infas) from the respondents' contact details. The individuals included are solely respondents who explicitly agreed to the linkage. People who had since withdrawn their willingness to participate in the panel and had previously agreed to the linkage were not included in the linkage. As a result, there is a difference between the number of people who at some time agreed to the linkage as part of the NEPS survey and the number of people available for linkage. In preparation for the data linkage, completely identical lines per person were deleted for the different survey waves. In the case of changes or differences in the spelling of the name or the address, multiple entries per person were retained as they could be relevant for the linkage.

To identify the NEPS respondents in the administrative data of the IAB, an excerpt from the central address files of the Data Warehouse of the Federal Employment Agency was used. These address files are restricted to the birth years surveyed in NEPS-SC4 and the postcode areas derived from the sampling points.

Both data sources may contain several different entries per person. Differences can arise, for example, due to relocations or name changes or as a result of actually identical information being recorded inconsistently. In order to achieve a high degree of linkage success, all the different details about individual persons are included in the linkage process.

4.3   Correction and standardisation of personal information

The above-mentioned personal information from NEPS and the administrative data was subjected to correction routines before the matching in order to achieve the best possible correspondence between the two data sources. For this so-called pre-processing, established procedures were used which were refined by the German Record Linkage Center (GRLC)[10] at the IAB-RDC for the peculiarities of the German context. See Christen (2012) for a comprehensive overview of all the methods described in chapters 4.3 to 4.5, and e.g. Schild & Antoni (2014) for further details on their adaptation to the German context.

Specifically, umlauts and special characters were replaced by their respective ASCII compliant equivalent, all letters were written as capitals, and redundant spaces and punctuation were removed. Several first or last names were combined into one box without spaces. Apart from these fundamental adjustments, variable-specific routines were performed in order to standardise the data. House numbers were extracted from a combined street/house number box and all other additions to addresses (e.g. floor, apartment no.) were deleted. Common patterns and abbreviations of street names were standardised. Concerning town names, annexes were standardised (e.g. towns with additional information like river or region) or deleted (e.g. Berlin-Kreuzberg).

4.4   Exact record linkage

After the name and address components had been corrected and standardised, the fields containing the entries of the two data sources were closely compared. First, all data lines were checked for an exact match of all available characteristics (first and last name, date of birth, gender, street, house number, postal code, town)[11]. The next step was to find out how many of the remaining (unlinked) people could be linked if a discrepancy in individual address fields or the date of birth was tolerated (while the first and last name fields still had to match exactly). The tolerance was chosen in such a way that the procedure entailed a very low risk of linking entries belonging to two different individuals (so-called false-positive matches). Therefore, discrepancies were first allowed in <u>one</u> of the particularly error-prone items: house number, postal code, street, town, year of birth or month of birth.

---

[10] For more information regarding the GRLC see Antoni & Schnell (2017) or http://www.record-linkage.de.

[11] A match concerns the corrected and standardised name and address fields, not the uncorrected characteristics. The pre-processing does not involve any corrections that would result in a loss of information in such a way that different individuals could be matched with each other.

4.5 Error-tolerant matching of the linkage variables

For some individuals it was not possible to find a match in the administrative address data using any of these exact matching procedures. In order to link these individuals with the administrative data, error-tolerant linkage procedures were used which compute similarity measures of the name and address components and select the highest quality row pairs. For each pair of variables to be matched, a similarity measure was calculated and summed up in a quality index. The calculations were performed using the software Merge Tool Box (MTB, see Schnell et al., 2004). For the first and last name as well as for the town and the streets a string similarity was calculated by means of the Jaro Winkler metric[12]. Concerning the house number, postal code, date of birth (broken down into the components day, month and year) and gender an exact comparison was made, because string similarity measures are not useful here.

In the error-tolerant linkage, the observations were grouped into blocks at the level of the three-digit postal code areas. This has the advantage that the number of comparisons required does not correspond to the Cartesian product of the lines presented in the two address files, but remains limited to the observations with an identical three-digit postal code. In addition, this reduces the risk of linking different people with very similar identifiers. However, it can happen to a limited extent that no correspondence can be found for individuals in the administrative data - for example due to missing or incorrect information about the postal code or because a move to another postal code area is only recorded in one of the two data sources. Thus, an additional linkage was performed using gender, year and month of birth as blocks.

A probabilistic matching procedure was used to calculate the similarity measure that arithmetically takes into account that the probability of a match of characteristics depends on their variance in the population. Here, so-called m- and u-probabilities are calculated, which indicate the probability of the personal details differing in the two different sources or how probable it is that an identical characteristic occurs in different individuals. For instance, a match in the gender variable does not indicate a real match to the same extent as a match of family names, and is therefore given a smaller weight. The variable-specific m- and u-probabilities underlying the probabilistic matching were determined in the context of past GRLC projects. All the available variables in the two data sources were compared and an additive similarity measure was calculated. The higher the value of this index, the greater the overall similarity of the two records is and the more likely it is that the match is genuine.

Of all the compared pairs of records, only those above a fixed threshold value of the similarity measure were considered as potential matches. Of these potential matches, for each respondent the corresponding value in the IEB data with the highest overall similarity was adopted. In other words, the best match was chosen for each person, if it was of sufficiently high matching quality. The threshold value was chosen in such a way that on the one hand the tolerance for discrepancies was sufficiently large so that actual matches were also identified as matches, but on the other hand not too small, as this would increase the likelihood of mismatches.

For the individuals matched by means of the error-tolerant linkage procedure, there are minor differences in the spelling of the fields first and last name, street or town. The smaller the differ-

---

[12] The n-grams algorithm was used to calculate the string similarity for the streets as a robustness check. Since the Jaro-Winkler algorithm found three additional matches while the remaining matches were identical, the Jaro-Winkler is being used.

ence, the higher is the calculated similarity measure. By regarding only observations with a measure of correspondence above a specified threshold value as potential matches, there was only a very small probability of different people being identified as a match within the error-tolerant linkage procedure. The respondents' gender and date of birth corresponded in 95% of the cases. In very rare cases either a digit in the date of birth or the information concerning the person's gender differed. However, a manual inspection revealed that these cases were plausible matches with incorrect information in the source material.

For the remaining individuals who were not matched by any of the mentioned methods a manual linkage was performed. To this end, the lines below the quality limit were manually inspected and plausible matches marked.

To verify the matches, various linkage steps were carried out and the resulting individual matches were compared, whereby the matches were found to be highly reliable. In addition, manual inspections were made on a random basis in order to check the plausibility of the matches. Especially the matches in which several details differed were checked.

4.6   Result of the data linkage

Up to wave 9 a total of 11,113 people gave their consent for the linkage (67.85% of all respondents). However, it was only possible to use 10,016 of these people for the data linkage. For legal reasons, individuals who had withdrawn their willingness to participate in the panel in the meantime were not available for the linkage, although they had previously agreed to it.

For 9,169 persons the survey data could be linked to an account in the administrative data. This corresponds to a very high linkage rate of 91.54% of the individuals available for the linkage between the NEPS survey data and the administrative data. Only 847 individuals (8.46%) could not be found in the administrative data, despite having given their consent for the linkage[13]. Table 7 summarises the success rates of the individual linkage steps.

Generated technical variables in the dataset (see chapter 6.2) indicate for each respondent which linkage method was used and to what degree of certainty a successful match was established. This allows users to decide at their own discretion which of the cases should be used in their own analyses.

---

[13] For 471 linked Persons administrative data was unavailable. Therefore, they are not included as matches in the following table. Due to their age, respondents more commonly did not enter the labour market yet. Thus, there is no administrative labour market data available for them.

Table 5 Results of the record linkage

	N	Percentage of those providing consent	Percentage of respondents
NEPS respondents	16,379		100.00%
Respondents providing consent for linkage	11,113		67.85%
Consenting respondents available for linkage	10,016	100.00%	61.15%
Matches with IEB	9,169	91.54%	55.98%
Including:			
Exact matches	8,735	87.21%	53.33%
Probabilistic matches	424	4.23%	2.59%
Manual matches	10	0.10%	0.06%

5 Data quality and problems

5.1 NEPS-SC4 survey data (NEPS-SC4-ADIAB)

A detailed description of the survey data, their structure and information on the sample and the survey in general can be found in the Data Manual of Starting Cohort 4, which also lists generated variables and details on weighting.[14]

5.2 Integrated Employment Biographies (IEB)

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 6 makes no claims to be exhaustive.

---

[14]See https://www.neps-data.de

Table 6 Biographical gaps and possible ways of identifying them

Biographical gap	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 3.1.6)	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)	

5.3 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 2.3) 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.[15]

- In contrast to SIAB 7514, SIAB 7517 contains the variable "Occupational status and working hours" (stib) again. However, the variable is only filled for reports that were submitted before the introduction of the new occupation code, and the FDZ does not extrapolate or impute for later reports. The categories of the variable stib that can be consistently observed over the whole observation period (occupational status as a trainee, distinction between part-time and full-time) are to be reproduced in the variables erwstat and teilzeit.

- The introduction of the new occupation code in 2011 led to a number of problems. For example, during the transition period granted to employers in the social security notification procedure,[16] 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 SIAB 7517. No imputation is performed regarding the gaps in the other variables.

- 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,

---

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

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

however, but at certain intervals. This is done using files of employment notifications for one particular year which were submitted 36, 18, 12 or 6 months after the end of the reporting year (e.g. the 18-month file for 2013 can be created in July 2015 at the earliest). Notifications submitted more than three years late are not taken into account at the IAB, which means that a 36-month file shows a 100 % degree of completeness by definition.

- In the version of the IEB on which NEPS-SC4-ADIAB is partly based, the year 2014 is the last year with a degree of completeness of BeH observations of 100%.[17] For the years 2015 and 2016, the 18-month files were used, and the observations for 2017 originate from a 6-month file. It can therefore be assumed that employment notifications for 2015 and 2016 are slightly underreported in the SIAB, and that those for 2017 are underreported to a slightly higher degree. However, this should not reduce the ability to analyse the data at individual level. The missing notifications occur more frequently in a few establishments, however. This means that in individual cases the establishment data, e.g. establishment size, are grossly incorrect and will change considerably in subsequent versions.

5.4   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 a fault in one magnetic tape, the benefit receipt data up to and including 1980 are only partially contained. Thus, in the present data product, too, it can be assumed that information on benefit receipt in that period is not available in full.

5.5   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 merely on the LHG sources.

- Longitudinal analyses of individuals are affected by inaccuracies as it is not possible to distinguish between changes in the benefit entitlement status and relocations into and out of districts whose institutions had problems delivering data.

- Also from 2007 onwards, cases of underrecording occur at times. These generally last one month and occur mainly in the authorised municipalities.

- Underrecording and overrecording occur in connection with changes in the type of institution responsible for implementing SGB II:
  – In the context of the reform of the territories covered by the institutions, which came into force on 1 January 2011, cases of underreporting occurred in the districts covered by the employment agencies of Dessau-Roßlau, Halberstadt, Halle and Sangerhausen.

---

[17] 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.

- Double notifications due to the territorial reforms in 2009/2011 and the changes in the form of the institutions as of 1 January 2012 and of 1 January 2013 are already corrected as far as possible in the IEB. Nonetheless double notifications may still occur.

- In the following job centres there are inaccuracies with regard to the allocation of benefit cases:
  - between Emden and Norden between September and December 2009
  - between Döbeln and Mittelsachsen from October to December 2012
  - between Tirschenreuth and Wunsiedel from November 2012 to March 2013

- Some individuals for whom a (X)LHG spell exists are excluded entirely or partly from benefit receipt according to SGB II, for instance because they take part in a subsidised training programme, receive an old-age pension, live in an in-patient facility or a residential institution or receive insurance payments aimed at avoiding need. This affects on average 3 to 5 percent of all cases. In XSozial this person group is sometimes underrecorded by some institutions. Exclusion from benefits cannot be identified in the SIAB.

- 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.

5.6 Jobseeker Histories (ASU/XASU)

5.6.1 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.

5.6.2 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.

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

5.7.1   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 Sections 2.5 and 2.7).

- The MTH only contains notifications that are recorded in BA procedures. The use of these procedures in cooperations of 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.

- Measures reported through the XSocial standard are not included in the MTH or NEPS-SC4-ADIAB.

- Because of the reorganisation of the institutions responsible for implementing SGB-II in 2011 to 2014, 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.[18] This might result in a split or a duplication of the spell of the measure.

- 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.

---

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

5.7.2 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.

- 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 in2012, 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 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 3.1.8).

- The figures or persons received in the XMTH from monthly key date counts differ in several respects from the published BA statistics.

6 Description of variables

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

6.1 Identifiers

6.1.1 Individual ID (ID_t)

Variable label	individual ID for NEPS-SC4-ADIAB
Variable name	ID_t
Category	identifiers
Origin	generated
Data type	numerical
Detailed description	The individual ID indicates which observations belong to the same person. It also indicates which persons correspond in survey and administrative data, i.e. ID_t is included in both survey and administrative data. This identification variable is system-independent, which is why it does not allow any conclusions to be drawn about the characteristics of the person or original identifiers.

6.1.2 Establishment ID (betnr)

Variable label	establishment ID for NEPS-SC4-ADIAB

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 economic class. If they do **not** belong to the same economic class, each branch office is regarded as a separate establishment and is given its own establishment number.
	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.

6.2   Generated technical variables

6.2.1   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.

6.2.2   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 Section 2).

6.2.3   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 for an example code with the „merge"-command in Stata 14.

6.3   Period of validity

6.3.1   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). 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 3.1.9.
	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 2.5 and 2.6). 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.

6.3.2   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). Since the notification logic might under certain circumstances per-mit 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 3.1.9.
	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 3.1.5 and 3.1.6). 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.

6.3.3   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).

6.3.4   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.)

6.4   Personal information

6.4.1   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.

6.4.2   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.

6.4.3   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.

6.4.4   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.
Anonymisation	Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (nation_gr) is provided.

6.4.5   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.

6.4.6   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.

6.4.7 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. **3) ASU, MTH** Until 30 June 2006, only up to nine children could be recorded. The value zero does not exist. For observations prior to 30 June 2006, the value zero was recoded to "missing", since it is not clear whether zero should be interpreted as "no children" or as "field not filled in". For observations after 30 June 2006, the variable is only recorded if children exist.

6.4.8 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

	11 University of applied sciences without further specifications
	12 University without further specifications

In notifications that rely on the new occupation code (see Section 5.3) 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
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

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
25   Other (promotion-) advanced training for graduates

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	ausbild-ung_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 gradu-ates

Notes on quality	**1) BeH**

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).

For the variable "Vocational training (imputed)" (ausbildung_imp, see Section 6.4.9), a method was applied to correct missing values or inconsistent changes of the training variable. However, this variable is only filled in the source BeH and has different categories than the variable ausbildung.

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.

6.4.9 Vocational training (imputed) (ausbildung_imp)

Variable label	vocational training (imputed)
Variable name	ausbildung_imp
Category	personal variables
Origin	BeH
Data type	numerical
Detailed description	The variable Vocational training (imputed) is a supplement to the variable Vocational training (ausbildung) and contains additional and harmonized information on the vocational training of employees for BeH spells. The variable thus offers a solution to the problems concerning the variable Vocational training (ausbildung) described in Section 6.4.8. The imputation procedure is described in Thomsen et al (2018), which is based on the work of Fitzenberger et al. (2016).
	As the variable only uses the training information from BeH notification and because the educational categories of the old and the new occupation codes had to be harmonized for the variable Vocational training (ausbildung), the variable ausbildung_imp has other categories than the variable ausbildung.

6.4.10 School leaving qualification (schule)

Variable label	school leaving qualification
Variable name	schule
Category	personal variables
Origin	BeH, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	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 2.3) 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
	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.

6.5   Information on employment, benefit receipt and job search

6.5.1   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
Detailed description	**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.
	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 6.5.12). 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.

6.5.2 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 5.3). 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.

6.5.3   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 5.3). 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
	D 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.

6.5.4   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 5.3). 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.

6.5.5   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 "Classifi-cation of Occupations 2010" (Klassifikation der Berufe 2010, KldB2010, Bun-desagentur 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 rec-orded 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 re-ported using the old occupation code 1988 (KldB 1988) (see Section 5.3). These val-ues are recoded to the KldB2010 by transferring the key area. As the new occupa-tion 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 occu-pation 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.

6.5.6 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 accumu-lation of missing values occurs in 1991 due to the reunification. No imputation is performed here.

6.5.7 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 2.3), 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.

6.5.8 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 voca-tional education and training', 'occupational status and working hours' and 'occu-pation' 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 distin-guished via the values 109 and 209. For employees in marginal part-time employ-ment, 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 ben-efit type. Thus, it is possible to distinguish whether a person receives unemploy-ment 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 pen-sion 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 anonymiza-tion procedure described in Section 3.1.9.
	4) ASU, XASU
	For ASU/XASU observations, the 'employment status' variable reports the job search status.
	A distinction is made between those who are "unemployed and seeking work", "not unemployed and seeking work", "seeking advice" and "not seeking work". Applicants who only want advice from the BA are considered as "seeking advice". In addition, there are rehabilitants and, before 2008, persons aged 58 or older who are not fully available for placement. As of 1 August 2016, this may also include persons who no longer require support, but who are nevertheless still under the job center's care.
	"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 proce-dure described in Section 3.1.9.
	5) MTH, XMTH
	For observations in the MTH and the XMTH, the 'employment status' variable indi-cates 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 5.6). 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.

6.5.9   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.

6.5.10   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%.

6.5.11   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%.

6.5.12 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.

6.5.13 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 SIAB. This information does not refer to the start date of the episode, but to the start of the original time period.

6.5.14 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

6.5.15 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.

6.5.16 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.

6.5.17 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.

6.5.18 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.

6.5.19 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 2.5. 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.

6.5.20 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

6.5.21 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. |

6.6   Location data

6.6.1   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.

6.6.2   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)

6.6.3   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 availa-

	ble 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.

6.6.4   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.

6.7   Establishment variables

6.7.1   Classification of economic activities 73 (w73_3)

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

6.7.2   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).

6.7.3   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.

6.7.4   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).

6.7.5 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.

6.7.6 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).

6.7.7 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.

6.7.8 w73_3 completed by extrapolation/imputation (w73_3_gen)

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

6.7.9 Type of imputation w73_3 (group_w73_3)

Variable label	type of imputation w73_3
Variable name	group_w73_3
Category	establishment variables
Origin	BHP
Data type	numerical
Detailed description	This variable indicates the type of completion for the w73_3_gen variable. It reports whether the respective value in w73_3_gen is consistent with the original

	value from w73_3, still missing / extrapolated or imputed based on recording tables.
	A detailed description of the procedure can be found in Eberle et al. (2011).

6.7.10   w93_3 completed by extrapolation/imputation (w93_3_gen)

Variable label	w93_3 completed by extrapolation/imputation
Variable name	w93_3_gen
Category	establishment variables
Origin	BHP
Data type	numerical
Hierarchy	section (1-digit code)	division (2-digit code)	group (3-digit code)	class (4-digit code)	sub-class (5 digit code) of economic activity
Detailed description	This variable indicates the economic activity as a 3-digit code in accordance with the WZ93 classification. From 1998 up to and including 2003, the variable contains the original values from w93_3. Before 1998 and after 2003, the information is either written back / continued or replaced with the help of recoding tables, so that the variable contains time-consistent information on the economic activity based on the economic activity classification WS93. A detailed description can be found in Eberle et al. (2011).
	Further information on the WS93 classification can be found in the description of variable w93_3.

6.7.11   Type of imputation w93_3 (group_w93_3)

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

6.7.12   w08_3 completed by extrapolation/imputation (w08_3_gen)

Variable label	w08_3 completed by extrapolation/imputation
Variable name	w08_3_gen
Category	establishment variables
Origin	BHP
Data type	Numerical
Hierarchy	section (1-digit code)	division (2-digit code)	group (3-digit code)	class (4-digit code)	sub-class (5-digit code) of economic activity
Detailed description	This variable indicates the economic activity as a 3-digit code in accordance with the WZ08 classification and is available from 2008 onwards. From 2008 up to and including 2017, the variable contains the original values from w08_3. Before 2008, the information is either written back or replaced with the help of recoding tables, so that the variable contains time-consistent information on the economic activity

	based on the economic activity classification WZ08. A detailed description can be found in Eberle et al. (2011).
	Further information on the WZ08 classification can be found in the description of variable w08_3.

6.7.13 Type of imputation w08_3 (group_w08_3)

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

6.7.14 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.

6.7.15 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).

6.7.16 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.

6.7.17 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).

6.7.18 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.

6.7.19 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.

6.7.20 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.

6.7.21 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.

6.7.22 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.

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

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

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

6.8   Linkage variables

6.8.1   Availability of administrative personal data (match_admin)

Variable label	availability of administrative personal data
Variable name	match_admin
Category	Linkage-relevant variable
Origin	generated
Data type	numerical
Detailed description	This variable is only included in the matching file. For every respondent in the NEPS-SC4 survey data, it displays whether observations from the administrative personal data are available for this person. In a few cases, this variable takes the value 0, although the variable match_typ indicates a successful match. In these cases, it was possible to achieve a match by comparing the identifiers of both data sources, but the IEB does not contain any episodes about that person.

6.8.2   Matching method (match_typ)

Variable label	matching type
Variable name	match_typ
Category	Linkage-relevant variable
Origin	Record linkage
Data type	numerical
Detailed description	This variable is only included in the matching file. It describes the stage of the matching process at which it was possible to identify a person in the administrative data of the IAB.

6.8.3   Quality of the match (quality)

Variable label	quality of the match
Variable name	quality
Category	Linkage-relevant variable
Origin	Record linkage
Data type	numerical
Detailed description	This variable is only included in the matching file. Reports a quality indicator for those individuals whose IAB data were merged using probabilistic methods. The higher the value of the indicator, the better the correspondence between the two datasets.

References

Antoni, M., Schmucker, A., Seth, S., Berge Philipp v. (2019). Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975-2017." FDZ-Datenreport 02/2019 (de), Nürnberg.

Antoni, M., Schnell, R. (2017). The past, present and future of the German Record Linkage Center (GRLC). In: Jahrbücher für Nationalökonomie und Statistik, online first.

Bender, S., Hilzendegen, J., Rohwer, G., & Rudolph, H. (1996). Die IAB-Beschäftigtenstichprobe 1975–1990. Beiträge zur Arbeitsmarkt-und Berufsforschung 197, Nürnberg.

Bertat, T., Dundler, A., Grimm, C., Kiewitt, J., Schomaker, C., Schridde, H., & Zemann, C. (2013). Neue Erhebungsinhalte „Arbeitszeit ", „ausgeübte Tätigkeit" sowie „Schul-und Berufsabschluss "in der Beschäftigungsstatistik. Methodenbericht der Statistik der Bundesagentur für Arbeit, URL: http://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Methodenberichte/Beschaeftigungsstatistik/Generische-Publikationen/Methodenbericht-Neue-Erbebungsinhalte-Arbeitszeita-ausgeuebte-Taetigkeit-sowie-Schul-und-Berufsabschluss-in-der-Beschaeftigungsstatistik.pdf (Zugriff: 05.04.2018).

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

Bundesagentur für Arbeit (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. Methodenbericht der Statistik der Bundesagentur für Arbeit. Nürnberg. URL: http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Wirtschaftszweige/Klassifikation-der-Wirtschaftszweige-1973-2003/Klassifikationen-der-Wirtschaftszweige-1973-2003-Nav.html (Zugriff: 05.04.2018)

Bundesagentur für Arbeit (2011). Klassifikation der Berufe 2010–Band 1: systematischer und alphabetischer Teil mit Erläuterungen. Klassifikation der Berufe 2010, Nürnberg. URL: http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB2010/Printausgabe-KldB2010/Printausgabe-KldB-2010-Nav.html (Zugriff: 05.04.2018)

Bundesanstalt für Arbeit (1988). Klassifikation der Berufe–Systematisches und Alphabetisches Verzeichnis der Berufsbenennung, Nürnberg. URL: http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB1975-1992/KldB1975-1992-Nav.html (Zugriff: 05.04.2018)

Christen, P. (2012). Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Berlin: Springer.

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

Eberle, J., Jacobebbinghaus, P., Ludsteck, J., & Witter, J. (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.

Fitzenberger, B., Osikominu, A., Völter, R. (2006). Imputation rules to improve the education variable in the IAB employment subsample. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg. 126, H. 3, S. 405-436.

Ludsteck, J., & Thomsen, U. (2016). Imputation of the working time information for the employment register data. FDZ-Methodenreport 01/2016 (en), Nürnberg.

Meinken, H., & Koch, I. (2004). BA-Beschäftigtenpanel 1998-2002. Codebuch, Nürnberg.

Paulus, W., & Matthes, B. (2013). Klassifikation der Berufe: Struktur, Codierung und Umsteigeschlüssel. FDZ-Methodenreport 08/2013 (de), Nürnberg.

Schild, C.-J., & Antoni, M. (2014). Linking survey data with administrative social security data–The project „Interactions Between Capabilities in Work and Private Life ".German RLC Working Paper No. wp-grlc-2014-02.

Schmucker, A., Eberle, J., Eberle, J., Ganzer, A., Steigmaier, J., & Umkehrer, M. (2018). Betriebs-Historik-Panel 1975-2016. FDZ-Datenreport 01/2018 (de), Nürnberg.

Schnell, R., Bachteler, T., & Bender, S. (2004). A toolbox for record linkage. Austrian Journal of Statistics, 33(1-2), 125-133.

Statistisches Bundesamt (2002). Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ93), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-1993.html (Zugriff: 12.06.2019).

Statistisches Bundesamt (2003). Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/GueterWirtschaftklassifikationen/Content75/KlassifikationWZ2003.html (Zugriff: 12.06.2019).

Statistisches Bundesamt (2008). Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html (Zugriff: 12.06.2019).

Statistisches Bundesamt (2019). Staats- und Gebietssystematik, Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluessel_pdf.pdf?__blob=publicationFile&v=5 (Zugriff: 12.06.2019).

Thomsen, U., Ludsteck, J., & Schmucker, A. (2018). Skilled or unskilled - Improving the information on qualification for employee data in the IAB Employee Biography. FDZ-Methodenreport, 09/2018 (en), Nürnberg.

Wermter, W., & Cramer, U. (1988). Wie hoch war der Beschäftigtenanstieg seit 1983? – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. Mitteilungen aus der Arbeitsmarkt-und Berufsforschung 21(4), 468-482.

7 Appendix

7.1 Frequency tables

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

7.2 List of abbreviations

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

IAB	Institut für Arbeitsmarkt- und Berufsforschung	Institute for Employment Research
IAB-FDZ	Forschungsdatenzentrum der Bundesagentur für Arbeit am IAB	Research Data Centre of the Federal Employment Agency at the IAB
IEB	Integrierte Erwerbsbiographien	Integrated Employment Biographies
ISIC	International Standard Industrial Classification of All Economic Activities	International Standard Industrial Classification of All Economic Activities
LeH	Leistungsempfängerhistorik	Benefit Recipient History
LHG	Leistungshistorik Grundsicherung	Unemployment Benefit II Recipient History
LIfBi	Leibniz Institut für Bildungsverläufe	Leibniz Institute for Educational Trajectories
LVA	Landesversicherungsanstalt	Land Social Insurance Office
MTH	Maßnahmeteilnehmer-Historik	Participants-in-Measures History File
NACE	Nomenclature génerale des activités économiques dans les communautés européennes	Nomenclature génerale des activités économiques dans les communautés européennes
NEPS	Nationales Bildungspanel	National Educational Panel Study
NEPS-SC4	NEPS Startkohorte 4	NEPS Starting Cohort 4
NUTS	Nomenclature des unités territoriales statistiques	Nomenclature des unités territoriales statistiques
SGB	Sozialgesetzbuch	German Social Code
SIAB	Stichprobe der Integrierten Arbeitsmarktbiographien	Sample of Integrated Labour Market Biographies
VerBIS	Vermittlungs- und Beratungsinformationssystems	Information System for Placement and Counselling
XASU	Arbeitsuchenden-Historik 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

Corresponding author
Nadine Bachbauer
Phone: +49 911 179-9579
Email: Nadine.Bachbauer2@iab.de