FDZ·Datenreport

Documentation of labour market data

NEPS–SC6 survey data linked to administrative data of the IAB (NEPS–SC6–ADIAB 7515)

Manfred Antoni,
Nadine Bachbauer,
Johanna Eberle,
Basha Vicari

Bundesagentur für Arbeit

NEPS-SC6 survey data linked to administrative data of the IAB (NEPS-SC6-ADIAB 7515)

Manfred Antoni, Nadine Bachbauer, Johanna Eberle, Basha Vicari
(Institute for Employment Research - Institut für Arbeitsmarkt- und Berufsforschung)

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

FDZ-Datenreporte (RDC data reports) describe RDC 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 6 „Bildung im Erwachsenenalter und lebenslanges Lernen" (NEPS-SC6) mit administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB).

Abstract

This data report describes the linked survey data of Starting Cohort 6 "Adult Education and Lifelong Learning" of the National Educational Panel Study (NEPS-SC6) with administrative data of the Institute for Employment Research (IAB).

Data availability

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

1 Introduction and short description

1.1 Introduction

This data report describes the data product "NEPS-SC6 survey data linked to administrative data of the IAB (NEPS-SC6-ADIAB 7515)", which is provided jointly by the Leibniz Institute for Educational Trajectories e.V. (LIfBi) and the Institute for Employment Research (IAB). The linked data consist of information gathered in interviews with participants of the Starting Cohort 6 (SC6) of the National Educational Panel Study (NEPS) and their personal data from the administrative data of the IAB. The administrative data are only provided if the respondents have consented to the data being linked and if they could be identified in the IAB data.

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

The content of the Starting Cohort 6 of the National Educational Panel Study is aimed at tracking educational and occupational profiles over the life course as well as detecting determinants of formal, non-formal and informal educational activities. The data permit analyses of adults' skills and their significance for the employment situation. In addition, the data can be used to examine the impact of the employment situation and the family situation on education choices, as well as the returns to formal qualifications, competences and work experience. Identifying opportunities and obstacles to learning and education is another focus of the NEPS-SC6 survey.

The survey of NEPS-SC6 includes the birth cohorts from 1944 to 1986. The sample consists of three subsamples. The IAB-ALWA study[2], which covers the birth cohorts from 1956 to 1986, the refreshment sample, as well as the pile-up sample, which adds the population of births from 1944 to 1955 to the study population[3].

---

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

[2] More detailed information regarding ALWA can be found in Kleinert et al. (2011) and for ALWA-ADIAB in Antoni & Seth (2012). NEPS-SC6-ADIAB replaces ALWA-ADIAB because the majority of the ALWA study is included in the NEPS. The advantage of ALWA-ADIAB is merely that ALWA recorded the complete location history since birth.

[3] The first wave of the NEPS Starting Cohort 6 is composed of information from the ALWA study, while the second wave consists of information from the first main survey in the NEPS study. In this report the wave logic of the NEPS data is used. By contrast, some documentation materials of the NEPS include the identification of NEPS main surveys (without consideration of the ALWA study).

The sample is based on a population-based data collection in 240 communities with 271 sample points. For the linked data product, only the download version of the NEPS data is available. The RemoteNEPS or on-site version, whose data are less strongly anonymised, cannot be deployed for NEPS-SC6-ADIAB.

The administrative data are part of the Integrated Employment Biographies (IEB) of the Institute for Employment Research (IAB). In the present case, they contain data on all individuals who were surveyed in the context of the NEPS Starting Cohort 6, who gave consent for the linkage of the data, and who exhibit one of the following states at least once during the observation period:

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

These data, which come from different sources, are merged and edited in the IEB and the statuses are depicted exact to the day.

Whilst the Employee History (Beschäftigtenhistorik - BeH) is the origin of the information on employment subject to social security and marginal part-time employment, the receipt of benefits in accordance with Social Code Book III (SGB III) and Social Code Book II (SGB II) is recorded in the Benefit Recipient History (Leistungsempfängerhistorik - LeH) and the Unemployment Benefit II Recipient Histories (Leistungshistoriken Grundsicherung - LHG and XLHG). The Jobseeker Histories (Arbeitsuchenden-Historiken - ASU and XASU) are the data source for the periods of job search recorded by the BA, whilst participation in employment and training measures is recorded in the Participants-in-Measures History File (Maßnahmeteilnahmehistorik - MTH).

In addition to these personal data, information on the employing companies can also be obtained from the Establishment History Panel (BHP) of the IAB (Schmucker et al., 2016).

The data report is structured as follows. Besides the introduction, Section 1 contains information on data access as well as an outline of the data, the volume structure and a list of variables of the administrative data. A description of the individual data sources can be found in Section 2. The editing of the data and the sampling procedure are explained in Section 3.

Sections 4 and 5 discuss the data linkage and the data quality, whilst the individual variables are described in Section 6.

1.2 Data use

1.2.1 Data access and data management

The NEPS-SC6-ADIAB data are weakly anonymised and may only be analysed in the context of on-site use at one of the locations of the Research Data Centre of the IAB and subsequent remote data access.

In order to be able to use the data, it is first necessary to submit an application to the Research Data Centre of the IAB (IAB-RDC)[4]. The Federal Ministry of Labour and Social Affairs (BMAS) makes the decision regarding the approval of the research project. When approval has been granted, a NEPS data use agreement is concluded with the Research Data Center of the LIfBi[5]. Finally, as soon as the data use agreement for the NEPS data is available to the IAB, a user contract is concluded between the researcher's institution and the IAB. Details on applying for the dataset and possibilities for data processing can be found on the homepage of the RDC at the IAB. Data access therefore only begins when the applicant has concluded contracts with both research data centres.

To answer questions about this data product, users can contact the staff of the research data centres at the IAB and the LIfBi. The division of work between the two bodies requires that users with questions about the NEPS data contact the LIfBi-RDC, whereas questions about the administrative data, technical aspects of data linkage or access will be answered by the IAB-RDC. For general questions about the NEPS, in addition to the LIfBi-RDC, the NEPS Forum[6] can also be consulted.

The survey data of the NEPS and the administrative data of the IAB are stored in separate files, which makes the data structure clear and saves data storage space. The personal data from the two data sources can be merged via the NEPS individual ID (ID_t), which is also contained in the administrative personal data.

A further element of NEPS-SC6-ADIAB is the Linkage File. It contains the individual ID as well as technical variables and provides information about whether the merge with the administrative data was successful and if so, to what extent. The file also includes individuals

---

[4] In order to differentiate between the two research data centres involved, the Research Data Centre of the Federal Employment Agency (BA) at the IAB is referred as the IAB-RDC and the Research Data Center of the LIfBi as LIfBi-RDC.

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

[6] See https://forum.neps-data.de/en-us/welcome.aspx.

from the NEPS-SC6 sample for whom no administrative data were found. It is therefore possible to perform selectivity analyses based on all the participants in the NEPS-SC6.

The administrative data, analogue to the NEPS data, include German as well as English labels[7]. These have a modular structure and are stored in several files following the structure of the SIAB. One module, which is henceforth called the Individual File, contains identifiers (individual IDs and establishment IDs), the personal variables, the information on employment, benefit receipt and job search activity, the variables regarding place of residence, and technical variables. These personal data include daily information for the period 1975 to 2014.

A second module, the Basis Establishment File, contains the establishment number, the year, information on the place of work and economic activities as well as further establishment-specific information as of 30 June of the years 1975 to 2014. For eastern Germany, the establishment data are available only from 1992 onwards. Thus, administrative establishment and personal data are stored in separate files (see Figure 1). In addition, the Basis Establishment File includes generated variables for classifications of economic activities that are consistent over time (see Eberle et al., 2011). In addition, further establishment variables and extension files of the BHP can be used, which are described in Schmucker et al. (2016). The administrative Individual and Establishment Files are linked via the establishment number and the year of the data record. The variables marked with an "*" in the list of variables (see page 18ff.) are contained in the Basis Establishment File.

---

[7] With the Stata codes *label language en* or *label language de* it is possible to switch to English or German labels, respectively.

Figure 1 Data storage of the administrative data
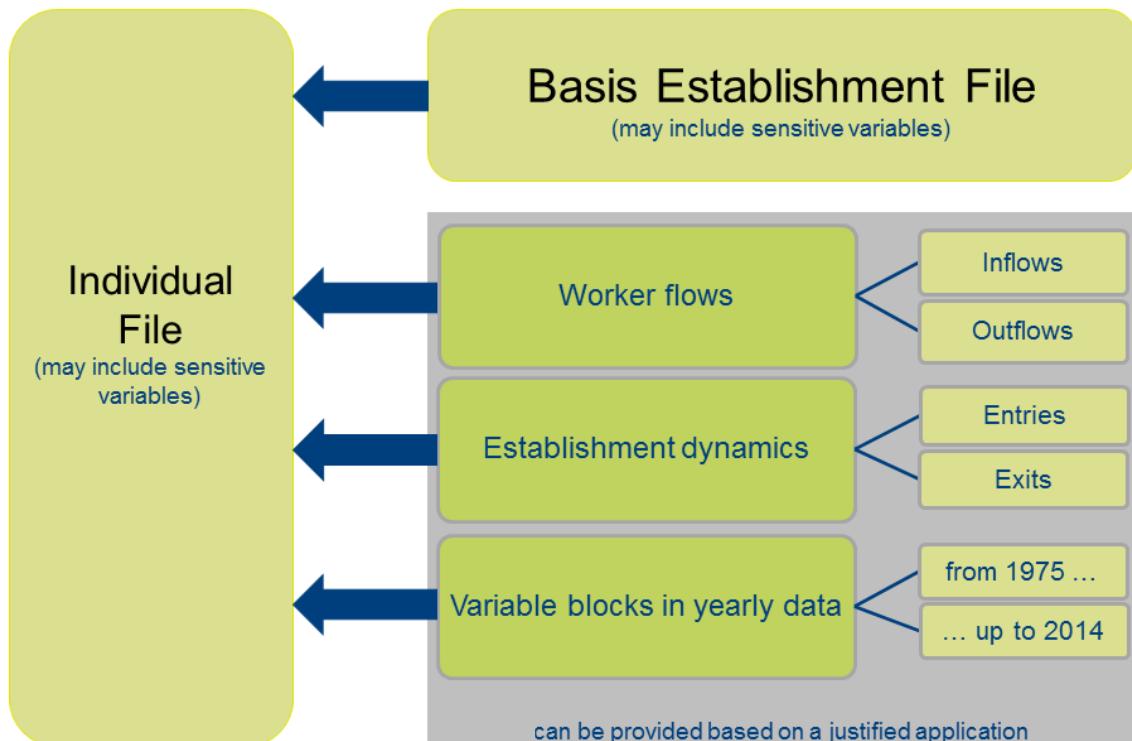
1.2.2 Sensitive variables and additional variables

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-SC6-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-SC6-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.

1.2.3 File names of the administrative data and the Linkage File

Individual File

NEPS-SC6-ADIAB_7515_v1.dta

Basis Establishment File

NEPS-SC6-ADIAB_7515_v1_bhp_basis_v1.dta

Linkage File

NEPS-SC6-ADIAB_7515_v1_schluessel.dta

Extension Files BHP

Core dataset with blocks of variables

NEPS-SC6-ADIAB_7515_v1_bhp_v1_*jjjj*.dta, *jjjj* = 1975 - 2014

Worker flows

NEPS-SC6-ADIAB_7515_v1_bhp_inflow_v1.dta

NEPS-SC6-ADIAB_7515_v1_bhp_outflow_v1.dta

Establishment dynamics

        NEPS-SC6-ADIAB_7515_v1_bhp_entry_v1.dta

        NEPS-SC6-ADIAB_7515_v1_bhp_exit_v1.dta

1.2.4 File names of the NEPS Starting Cohort 6

Table 1 Data structure NEPS-SC6

Dataset	File name	Content and special features
Basic data	Basics.dta	Information on current basic characteristics of the respondents (e.g. sociodemography, current employment and household characteristics); cross-sectional data in wide format
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	Methods.dta	Information on the field characteristics and data collection (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
Survey data	pTarget.dta	Information from the interviews with the target person; longitudinal data in long format
Competence data	xTarget Competencies.dta	Information on the competence tests carried out in waves 3, 5 and 7; longitudinal data in wide format
Methods data of the competence test	Methods Competencies.dta	Information on the methodical implementation of competence tests; Longitudinal data in long 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-centred episode spells
Marital state spell data	MaritalStates.dta	Information on events related to marital status (married, divorced and widowed); dataset generated from spPartner
Children data	Children.dta	Basic information on children with whom the target person has lived together at any time; dataset generated from spChild
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
Further education data	Further Education.dta	Integrated information on all course data, also retrospectively; dataset generated from the spell records for further education and training
School education spell data	spSchool.dta	Information on the general education history, from school enrolment to the (anticipated) time of completion; regional data (country, federal state)
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); regional data (country, federal state)
Employment spell data	spEmp.dta	Information on regular work episodes including internships; beginning of a new episode in the event

Dataset	File name	Content and special features
		of a change of employer, employment or career break; regional data (country, federal state)
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
Partner spell data	spPartner.dta	Information on the history of the target person's partnerships (cohabitation, marriage, separation, divorce); characteristics of the partner (e.g. date of birth, education, occupation)
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
Gap spell data	spGap.dta	Information on gaps in the recording of the individual life course; identified by test module
Occupational courses 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
Further courses spell data	spFurtherEdu1.dta	Information on courses completed in addition to the courses reported in spCourses and spVocTrain; both professional and private courses (e.g. yoga, cookery classes)
Additional course information spell data	spFurtherEdu2.dta	Additional information on two randomly selected courses from spVocTrain, spCourses and spFurtherEdu1 (e.g. costs, motivation, certificates)
German courses spell data	spFurtherEdu3.dta	Information about courses in German as a foreign language; surveyed only for migrants
Foreign school qualifications spell data	spSchoolExtExam.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	spVolunteerWork.dta	Information on voluntary activities (e.g. duration, requirements, content of the activity)
Residence spell data	spResidence.dta	Housing biography information (for original ALWA respondents only); regional data (country, federal state)

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

Table 2 Profile of NEPS-SC6-ADIAB

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 (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:
	Persons living in Germany (at the time of sampling) of working age regardless of employment status, nationality and German language skills
	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 (ALWA): 6,778 participants (linked: 5,354 persons)
	Wave 2: 11,649 participants (additionally linked: 3,416)
	Wave 3: 9,336 participants
	Wave 4: 14,112 participants (additionally linked: 3,888)
	Wave 5: 11,696 participants
	Wave 6: 10,639 participants
	Wave 7: 9,770 participants
	Wave 8: 9,236 participants
	Administrative data:
	12,660 persons
	454,974 original records
	599,833 records without any overlaps (after episode splitting)
Period	Survey data:
	wave 1 (ALWA): 2007/2008
	wave 2: 2009/2010
	wave 3: 2010/2011
	wave 4: 2011/2012
	wave 5: 2012/2013
	wave 6: 2013
	wave 7: 2014/2015
	wave 8: 2015/2016
	Administrative data:
	The period depends on the data source.
	BeH: 01.01.1975 - 31.12.2014 (2012: 30-month file, 2013: 18-month file; 2014: 6-month file)
	LeH: 29.11.1976 - 31.12.2014
	ASU: 07.11.1994 - 31.12.2014
	LHG: 01.01.2005 - 31.12.2014
	XASU: 01.01.2005 - 31.12.2014
	MTH: 02.11.1993 - 31.12.2014
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 31.12.2014
Survey design	Survey data:
	Wave 1: ALWA (Working and Learning in a Changing World)
	Wave 2: Follow-up ALWA survey, refreshment of as well as addition to the panel, CATI-CAPI biographical interview
	Wave 3: Follow-up survey previous wave, CATI-CAPI biographical interview and PAPI test of competence (reading/maths/metacognition)
	Wave 4: Follow-up survey previous wave and refreshment, CATI-CAPI biographical interview

	Wave 5: Follow-up survey previous wave, CATI-CAPI biographical interview and PAPI test of competence (reading/natural sciences/ICT/metacognition)
	Wave 6: Follow-up survey previous wave, CATI-CAPI biographical interview
	Wave 7: Follow-up survey previous wave, CATI-CAPI biographical interview and PAPI test of competence (basic cognitive skills/listening comprehension/metacognition)
	Wave 8: Follow-up survey previous wave, CATI-CAPI biographical interview
	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; 31 single data files between 66 KB and 143.1 MB Administrative data: Stata; Personal data set: 42.4 MB; Establishment data set: 1.65 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	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-SC6 survey data linked to administrative data of the IAB (NEPS-SC6-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-SC6-ADIAB7515.de.en.v1

	Data documentation: Antoni, Manfred; Bachbauer, Nadine; Eberle Johanna, Vicari, Basha (2018): NEPS-SC6 survey data linked to administrative data of the IAB (NEPS-SC6-ADIAB 7515). FDZ-Datenreport, 02/2018 (en), Nürnberg. DOI: 10.5164/IAB.FDZD.1802.en.v1
Dataset version	NEPS-SC6 survey data linked to administrative data of the IAB (NEPS-SC6-ADIAB 7515); DOI: 10.5164/IAB.NEPS-SC6-ADIAB7515.de.en.v1

1.4 List of variables in the administrative data

The overview of variables in Table 4 lists the variable names and the longer descriptions of the variables. 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)

Table 4 also shows which variables are available for which data sources, and Table 3 shows how to interpret the degree of completeness.

Table 3 Degree of completeness of the variables

	Variable is available for the source. Degree of completeness always > 0.85
	Variable is available for the source. Lower or changing degree of completeness, see variable description and frequency count
	Variable is not available for this source. Degree of completeness always < 0.05

Example: the variable 'daily wage, daily benefit rate' 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 observations it is the measure category. These differences are not immediately obvious from the variable name for every variable.

Table 4 List of variables in the administrative data and their degree of completeness

List of variables	BHP Basis	Page	BeH	LeH	LHG	ASU	XASU	MTH
Identifiers		**44**						
Person-ID of the NEPS-SC6-ADIAB (ID_t)		44						
Company-ID of the NEPS-SC6-ADIAB (betnr)	*	44						
Generated technical variables		**44**						
Observation counter per person (spell)		45						
Source of spell (quelle)		45						
Year (jahr)	*	46						
Validity period		**46**						
Original start date of observation (begorig)		46						
Original end date of observation (endorig)		47						
Start date of split episode (begepi)		47						
End date of split episode (endepi)		48						
Personal information		**48**						
Gender (frau)		48						
Year of birth (gebjahr)		48						
Nationality (nation)		48						
Nationality, aggregated (nation_gr)		49						
Marital status (famst)		49						
Number of children (kind)		49						
Vocational training (ausbildung)		50						
School leaving qualification (schule)		51						
Information on employment, benefit receipt and job search		**53**						
Daily wage, daily benefit rate (tentgelt)		53						
Occupation – current/most recent (KldB 1988) (beruf)		54						
Occupational group – current/most recent (KldB 2010) (beruf2010_3)		55						
Occupational sub-group – current/most recent (KldB 2010) (beruf2010_4)		55						
Level of requirement – current/most recent (KldB 2010) (niveau)		56						
Part-time (teilzeit)		57						
Employment status (erwstat)		58						
Transition zone (gleitz)		59						

List of variables	BHP Basis	Page	BeH	LeH	LHG	ASU	XASU	MTH
Temporary agency work (leih)		60	▫					
Fixed-term contract (befrist)		60	▫					
Reason for submitting the employment notification/for the end of benefit receipt (grund)		60	▪	▪	▫	▪	▫	
Employment status prior to job search (estatvor)		62				▪	▫	
Employment status after job search (estatnach)		62				▫		
Client profile (profil)		63				▫		▫
Type of termination of last job (art_kuend)		63				▫		▫
Desired working hours of the job sought (arbzeit)		64				▫		▫
Residual claim/planned duration (restanspruch)		64		▪				▪
Type of institution (traeger)		65			▪	▪	▪	▪
Start date of unemployment (alo_beg)		65	▫	▫	▫	▫	▫	▫
Duration of unemployment (alo_dau)		65	▪	▪	▪	▪	▪	▪
Location information		**66**						
Place of residence: district (Kreis) (wo_kreis)		66						
Place of residence: federal state (Bundesland) (wo_bula)		67	▪	▪	▪	▪	▪	▪
Place of residence – employment agency (wo_aa)		67	▪	▪	▪	▪	▪	▪
Place of residence: regional directorate (wo_rd)		68	▪	▪	▪	▪	▪	▪
Establishment variables		**68**						
Economic activity 73, 3-digit code (w73_3)	*	68	▫					
Economic activity 93, 5-digit code (w93_5)	*	69	▫					
Economic activity 93, 3-digit code (w93_3)	*	69	▫					
Economic activity 03, 5-digit code (w03_5)	*	70	▫					
Economic activity 03, 3-digit code (w03_3)	*	70	▫					
Economic activity 08. 5-digit code (w08_5)	*	71	▫					
Economic activity 08, 3-digit code (w08_3)	*	71	▫					
Economic activity 73 generated – completed by extrapolation/imputation (w73_3_gen)	*	72	▪					

List of variables	BHP Basis	Page	BeH	LeH	LHG	ASU	XASU	MTH
Economic activity 73 generated – type of completion (group_w73_3)	*	72						
Economic activity 93 generated – completed by extrapolation/imputation (w93_3_gen)	*	73						
Type of completion w93_3 (group_w93_3)	*	73						
Year of first appearance of establishment number (grd_jahr)	*	73						
First appearance of establishment number (grd_dat)	*	74						
Year of last appearance of establishment number (lzt_jahr)	*	74						
Last appearance of establishment number (lzt_dat)	*	75						
Total number of employees (az_ges)	*	75						
Number of full-time employees (regular workers + others) (az_vz)	*	75						
Number of employees in marginal part-time employment (az_gf)	*	76						
Mean imputed wage all full-time employees (te_imp_mw)	*	76						
Place of work: district (ao_kreis)	*	77						
Place of work: federal state (Bundesland) (ao_bula)	*	77						

1.5 Volume structure

Table 5 shows the number of episodes by source and before and after episode splitting, which is discussed in more detail in Chapter 3.2.

Table 5 Volume structure

No. of cases	Before splitting	After splitting
BeH	344,741	396,736
LeH	35,641	55,889
LHG	9,965	30,257
ASU	52,641	94,632
XASU	2,286	4,872
MTH	9,700	17,447
Total records	454,974	599,833
Persons		12,660

2 Data sources

2.1 NEPS-SC6 survey data (NEPS-SC6)

The NEPS-SC6 sub-study "Adult Education and Lifelong Learning" provides a data basis which allows research on education, employment and skills development throughout the life of working-age adults, which means from the age of 23 to retirement and beyond. The aim of this sub-study is to examine adult education and to close the existing knowledge gaps regarding adult education as well as skills and changes in them after initial education and training. To this end, the full range of educational activities and learning processes (formal, non-formal and informal education), decisions leading to participation in such activities, the tasks at the workplace and the life courses of the respondents are recorded in detail. In addition, information is also gathered about reading, mathematics, science and ICT literacy skills as well as non-cognitive skills such as personality, motivation and social skills.

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

- to trace the educational trajectories of adults beyond their employment history and beyond retirement age, and to track the education and employment histories of younger cohorts after they have entered the workforce;
- to identify the determinants of decisions to participate in formal or non-formal learning activities after initial training;
- to describe the competences of different groups of adults in Germany and to explain competence development in adulthood and the role that employment plays in it;
- to analyse the effects of specific education contexts in adulthood, especially the employment situation and the family constellation, on education decisions and participation in further education and training;
- to estimate the returns to formal qualifications, competencies and work experience in the form of wages, occupational careers and elements of other areas of life such as well-being or social commitment;
- to generate empirical results on the competencies of individuals with a migration background, their resources, their participation in continuing education and training and its returns;
- to identify opportunities and obstacles to learning and education in later adult life.

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 and were edited in terms of the selection and characteristics of variables along

the same lines as the Sample of Integrated Labour Market Biographies (SIAB, see Antoni et al., 2016).

The IEB unite data from five 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-SC6-ADIAB as well as their relationship to other RDC data products.

Figure 2 Sources of the NEPS-SC6-ADIAB 7515
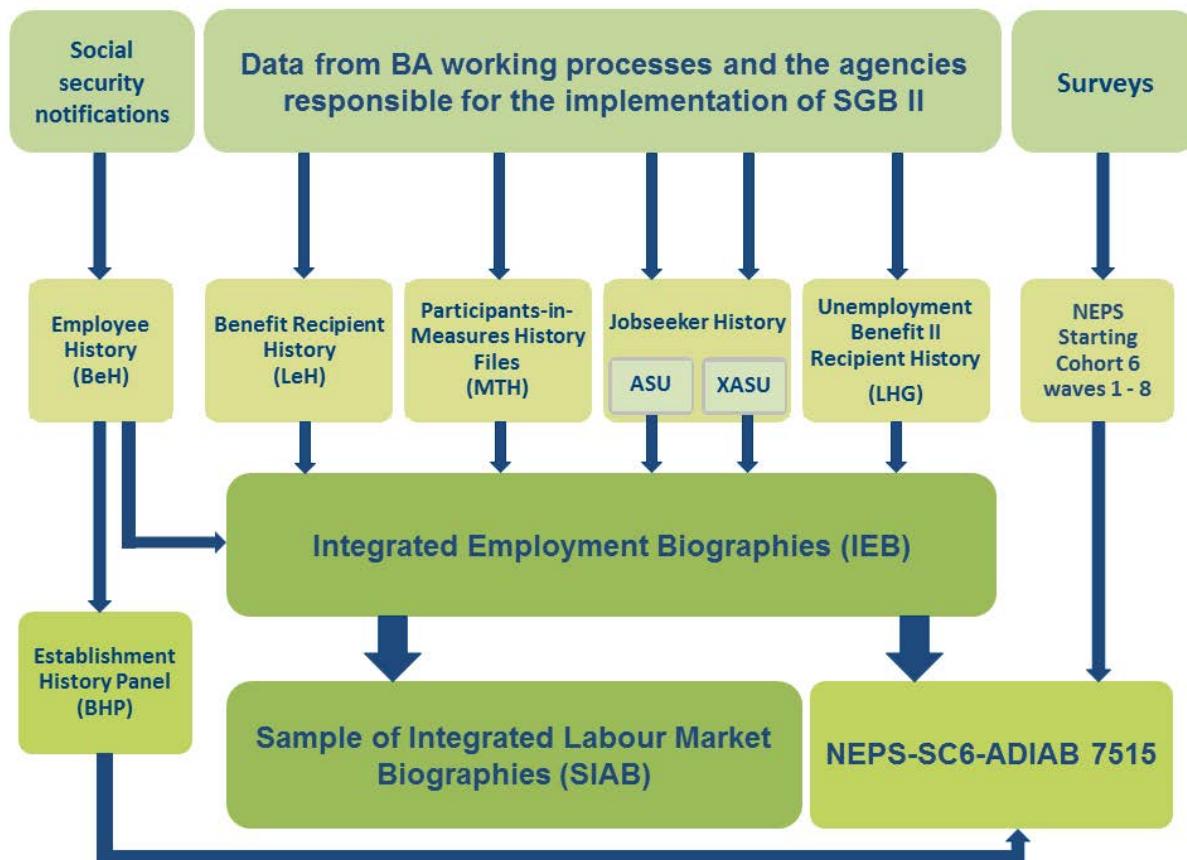
2.3 Employee History (BeH)

The source of data regarding employment is the Employee History (Beschäftigten-Historik - 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 eastern Germany as of 1 January 1991) and is known by the abbreviation DEÜV (previously DEVO/DÜVO) (for further details see: Bender et al. 1996, p. 4 ff.; Wermter & Cramer 1988). Under this procedure employers are required to submit notifications to the

responsible social security agencies concerning all of their employees covered by social security 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[8] (see Cramer 1985) are in principle not recorded in the BeH. Since the notification procedure was changed on 1 January 1999, employees in marginal part-time employment and unpaid family workers have also been recorded (not contained in the data until 1 April 1999). The data are recorded by the health insurance companies, collected 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 it has to be taken 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[9].

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

Not all the BHP variables are included as standard in NEPS-SC6-ADIAB. Upon request, individual additional BHP variables[10] may be merged by the RDC.

```
use NEPS-SC6-ADIAB_7515_v1.dta, clear
gen int jahr = year(begepi)
sort betnr jahr
merge m:1 betnr jahr using NEPS-SC6-ADIAB_7515_v1_bhp_basis_v1.dta
```

Box 1 Sample code for Stata

---

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

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

[10]Further information regarding the BHP can be found in Schmucker et al. (2016) and under http://fdz.iab.de/

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 and maintenance allowance, in other words not benefits under the sphere of Social Code Book II (e.g. unemployment benefit II). 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 benefit and capable of work, about the members of their benefit community (Bedarfsgemeinschaft) in accordance with § 7 SGB II and about certain individuals associated with the benefit community. In the SIAB it is not possible, however, to link individuals with benefit receipt in accordance with Social Code Book II (SGB II) at the level of benefit communities. 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 community in the case of Unemployment Benefit II, it is difficult to assign an individual benefit rate and this is not done in the SIAB.

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) / municipalities exercising their duties separately (until 2011) – here the tasks are divided between the BA and the municipality [11],

---

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

– authorised municipalities, which are also called opting local authorities or opting municipalities according to the initial experimental clause of Section 6a SGB II - 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[12]. 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. 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 5.5).

2.6 Jobseeker History (ASU und XASU)

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

2.7 Participants-In-Measures History Files (MTH)

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. This means in particular that no measures implemented by the authorised municipalities are recorded in the MTH as these are reported via a different standard, XSozial. Information from these institutions is not included in the IEB due to a number of data problems. The earliest available data in the MTH are from 1 January 2000.

---

[12] 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-SC6 survey data

Detailed information on sampling[13], data collection, the instruments used, and the structure and processing of NEPS-SC6 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 SC6 including variable names, variables and value labels (but no observations)[14].

3.1.2 Administrative data from the Integrated Employment Biographies (IEB)

Before the data from the data 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.

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

Unlike in the IAB Employment Sample, 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.

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

---

[13] For the sampling method of the NEPS see Hammon et al. (2016).
[14] 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 earnings amounting to zero or 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 2014.

3.1.4 Benefit Recipient History (LeH)

- Observations without a valid start date are excluded.
- Observations whose end date precedes the 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 for the BeH

3.1.5 Unemployment Benefit II Recipient History (LHG)

- Observations without a BA client number are deleted.
- Observations without a valid date of birth 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 community (BG) are depicted. 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 community (14, 15, 18 and 25) and the individual retirement age of members of the BG (see Section 3.1.7),
    - when the structure of the benefit community changes (e.g. due to entries/exits),
    - when there are changes in a variable of the BG client and

- o 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 were corrected as far as possible.

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

3.1.6 Jobseeker History (ASU and XASU)

- Observations whose end date precedes the start date are not included in the ASU.

- 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 ASU data basis only distinguishes between observations with the status "unemployed" and "jobseeker", and since 2006 "seeking advice" and "without status". In the IEB, however, the additional status "ill/not able to work" is available. The employment status "ill/not able to work" is assigned to IEB spells when in the ASU data basis

- o a preceding observation with the status "unemployed" exists which joins the next observation without a gap and has "incapacitated for work" as the reason for exit, and

- o a subsequent observation with the status "unemployed" exists which also follows without a gap, and

- o the observation itself does not have the status "unemployed" but "jobseeker".

  - In contrast to the ASU source, the XASU only distinguishes between the status
    - o "not unemployed, but seeking work" or
    - o "unemployed and simultaneously seeking work".

  - Unlike in the ASU, periods of illness are not taken into account when generating the "employment status", since no information about illness is available in the XASU data.

When calculating the unemployment duration with XASU observations, gaps due to illness cannot be identified.

− The XASU contains non-overlapping time periods for individuals. If one of the following variables changes, in each case a new data spell is generated for the XASU:

  o change of job search status

  o change of availability

  o change of SGB II institution (due to notification procedure)

  o change of place of residence

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

3.1.7 Participants-In-Measures History Files (MTH)

− Observations whose end date precedes the start date are excluded.

− 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.

3.1.8 SGB-II anonymisation

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

In order to prevent an indirect determination of the exact 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 "employment status" variable.

3.2 Episode splitting

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

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

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

Figure 3 Episode splitting
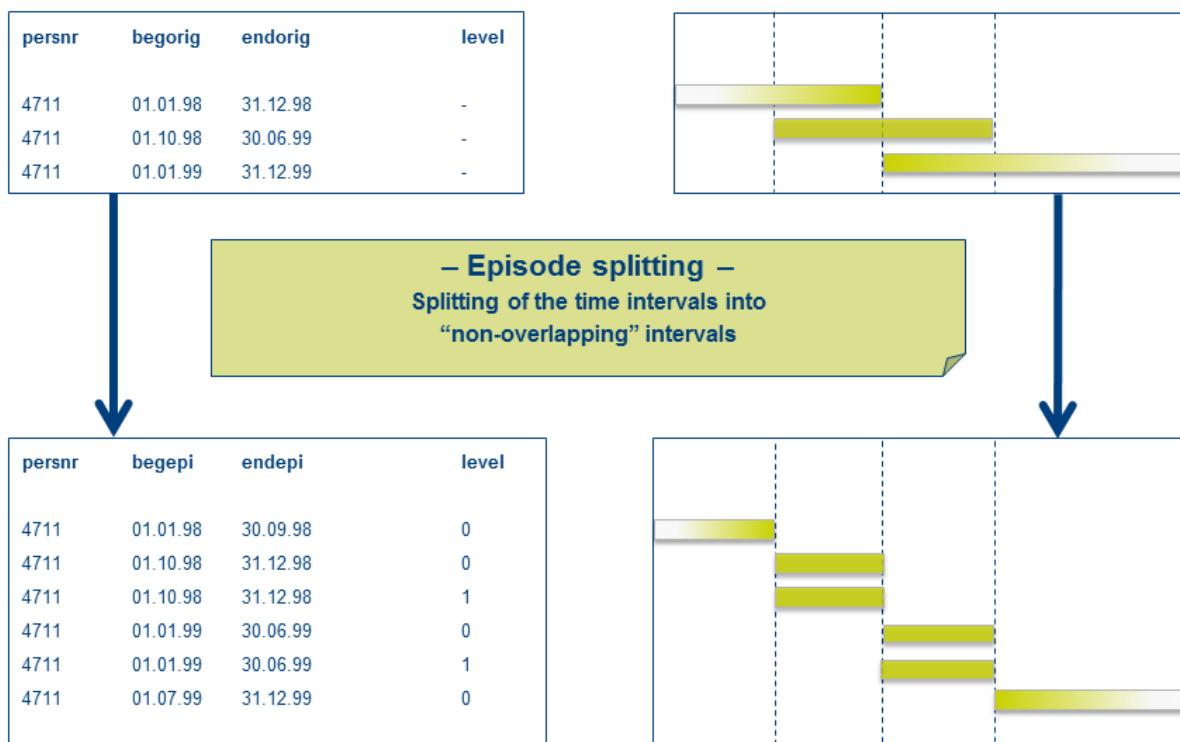
It is advisable to sort entirely parallel observations generated by the splitting procedure in a consistent manner. The variables 'observation counter per episode' (level2) and 'observation

counter per episode and source' (level1) that were previously contained in the SIAB can be generated using the following Stata commands if required:

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

Box 2 Stata code for the generation of counter per episode

3.3 Missing values

In the data missing values are coded as follows (Table 6):

Table 6 Coding of missing values

Term	Value	Description
No (valid) details available	.z	Values of a variable which are not systematically missing, i.e. the variable is available in principle for the data source, but no details are available for the value considered or cannot be interpreted reasonably.
Systematically not available	.n	A variable is not available in principle for a data source (dark grey cells in the overview of variables in Section 1.4) 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-SC6. 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 1 through 6, 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-SC6 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. Prior to this, a restriction was made to the birth years surveyed in NEPS-SC6 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 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)[15] 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)[16]. The next step was to find out how many more 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 *one* of the particularly error-prone items: house number, postal code, street, town, year of birth or birthday/month. As a last step in the exact matching process, people were linked if the first and last name, gender, date of birth and the first three digits of the postal code corresponded but there was a discrepancy in the fields street, house number, town and the last two digits of the postal code. In this way it should be possible to identify people who had moved within a certain radius (and that move was registered only in one of the two sources). The restriction to the three-digit postal code areas was made against the background that in

---

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

[16] 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.

this way the risk of false matches was kept low, since the probability of a perfect match of full name and date of birth within such regional boundaries is very low.

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 a string similarity was calculated by means of the Jaro Winkler metric, for streets the n-gram algorithm was used[17]. 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. For comparison, an additional linkage was performed without blocking at the level of three-digit postal codes. No significant differences were found in the results. The reason is that in practice, postal code information is seldom flawed, especially not in the first three digits. In addition, changes of address were well recorded both in the administrative address records and in the respondent addresses supplied by the infas survey institute.

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

---

[17] With the Jaro-Winkler algorithm, similarities at the beginning of a word are weighted comparatively more highly, whereas with n-grams, all bigrams go into the calculated string similarity with equal weights. While name and place fields often contain additional information that make a higher weighting of the beginning of the word seem meaningful, n-grams can be assumed to be comparatively more suitable in the case of street names.

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, provided it was of sufficiently high 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 difference, 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 5 a total of 15,982 people gave their consent for the linkage (93.24% of all respondents). However, it was only possible to use 14,065 of these people for the data linkage, as 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 12,660 persons, the survey data could be linked to an account in the administrative data. This corresponds to a very high linkage rate of 90.1% of the individuals available for the linkage between the NEPS survey data and the administrative data. Only 1,405 individuals (9.99%) could not be found in the administrative data, despite having given their consent for the linkage. 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 with what degree of certainty a successful match was established. This allows each user to decide at their own discretion which of the cases should be used in their own analyses.

Table 6 Results of the record linkage

	N	Percentage of those providing consent	Percentage of respondents
NEPS respondents	17,140		100.00%
Respondents providing consent for linkage	15,982	100.00%	93.24%
Consenting respondents available for linkage	14,065	100.00%	82.06%
Matches with IEB	12,660	90.01%	73.86%
including:			
Exact matches	11,632	82.70%	67.86%
Probabilistic matches	1,012	7.20%	5.90%
Manual matches	16	0.11%	0.09%

5 Data quality and problems

5.1 NEPS-SC6 survey data (NEPS-SC6)

A precise description of the data, its structure as well as information about the sample and the questionnaire in general can be found in the Data Manual of Starting Cohort 6. It also lists the generated variables and details about the weighting[18].

---

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

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 recognise 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 'reason for notification / reason for end of benefit receipt / reason for discontinuation of SGB II / reason for deregistration' and 'employment status' variables in the various sources (see Table 8). The list makes no claims to be exhaustive.

Table 7 Biographical gaps and identification possibilities

Biographical gap	Information on gap, possibly identifiable using the details in the "grund" variable in the preceding observation of the source
Civil servants, professional soldiers, judges, employees of bodies or foundations under public law	XASU
Self-employed persons without support	LeH, ASU
Students, persons in school-based further education	LeH, LHG, ASU, XASU
Persons who are ill/not able to work for more than 6 weeks (illness during unemployment, however, is represented in the ASU source under certain circumstances, see Section 3.1.6).	BeH, LeH, ASU
Persons receiving old-age pension without employment if not a member of a benefit community	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 completing civilian service	BeH, LeH, LHG, ASU, XASU
Persons completing a year of voluntary social or ecological work instead of civilian 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)	
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)

− 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,[19] there was a temporary increase in the number of missing details. Analyses of the BA statistics (Bertat et al., 2013) show that in 20 to 30 percent of cases no information was contained in the new or converted variables 'occupation - activity performed', 'working hours' and 'vocational education and training' after the switchover. 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 used in the NEPS-SC6-ADIAB 7515.

---

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

- Due to the introduction of the employment notification procedure in the federal states of eastern Germany, the notifications for eastern Germany can only be assumed to be sufficiently complete from 1993 onwards. For the same reason, a large number of spells for 1991 have missing values for several variables (such as employment status).

- The increase in the number of BeH observations from 1999 onwards is due to the introduction of the obligation to submit employment notifications for people in marginal part-time employment from 1 April 1999 onwards.

- Especially in 1999, observations of part-time employment increase significantly. This is caused by the actually observed increase in part-time work as well as by the fact that since 1999 employment notifications have been completed more correctly.

- Within the employment notification procedure, a certain time lag is unavoidable. Although changes in employment relationships have to be reported immediately, and existing employment relationships have to be confirmed annually by April (until the annual notification 2012) or mid-February (from the annual notification 2013 onwards) of the following year, some notifications actually arrive years later. The History File of the IAB is not updated continuously, however, but at certain intervals. This is done using files of employment notifications for one particular year which were submitted 36, 18, 12 or 6 months after the end of the reporting year (e.g. the 18-month file for 2013 can be created in July 2015 at the earliest). Notifications submitted more than three years late are not taken into account at the IAB, which means that a 36-month file shows a 100 % degree of completeness by definition.

- In the version of the IEB on which the SIAB data are based the degree of completeness of the BeH observations last stood at 100 % in 2012, since only the 30-month material could be used for this year.[20] 18-month files were used for 2013, and the observations for 2014 come from a 6-month file. It can therefore be assumed that the BeH observations for 2013 and in particular for 2014 are slightly underreported in the SIAB. However, this should not reduce the ability to analyse the data at individual level. The missing notifications occur more frequently in a few establishments, however. This means that in individual cases the establishment data, e.g. establishment size, are grossly incorrect and will change considerably in subsequent versions.

- In 1984, a change was made in the employment notification procedure. From that time onwards, one-off payments of gross earned income were reported as part of the annual

---

[20] Due to a redesign of the BeH data basis, only the 30-month instead of the 36-month material was available at the time the BeH was created. Analyses using previous data versions have shown, however, that the 18-month file generally already displays a degree of completeness of around 99 percent. Therefore, there are very few notifications of subsequent information or corrections during the following 18 months.

earnings subject to social security contributions, which leads to an increase in the average daily wage. In particular, the proportion of wages and salaries above the upper earnings limit increased considerably from that year onwards (see Bender et al. 1996).

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

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

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

− In the years 1996 to 1998, the values 841-844 (doctors and pharmacies) within the 'occupation – activity performed' variable are very rare compared to the adjacent years.

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.

− Due to an internal change of systems, there is a break in the recording of periods of exclusion from benefits and of benefit suspension in 2004. Until 1 July 2004 periods of exclusion from benefits and of benefit suspension can only be identified via the 'reason for end of benefit receipt' in the preceding LeH observation. After this date a separate observation is available with the daily benefit rate = 0 for periods of benefit exclusion and suspension.

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 under-recording occur at times. These generally last one month and occur mainly in the LHG.
- Under-recording and over-recording occur in connection with changes in the type of institution responsible for implementing SGB II:
  o 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.
  o Double notifications due to the territorial reforms in 2009/2011 and the changes in the form of the institutions as of 1 January 2012 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:
  o between Emden and Norden between September and December 2009
  o between Döbeln and Mittelsachsen from October to December 2012
  o 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 under-recorded by some institutions. Exclusion from benefits cannot be identified in the NEPS-SC6-ADIAB.

5.6 Jobseeker History (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.
- 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 'type of institution' variable 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 to mid-2006, the computerised procedure coArb, from which the jobseeker data originated, was replaced by the VerBIS procedure at the BA. From July 2005, coArb was first replaced by VerBIS as a pilot project at the Wiesbaden agency, and then successively in several instalments in all other employment agencies from December 2005. Many of the recorded variables were surveyed in the two systems with varying degrees of differentiation and qualitative weighting. Integrating these variables into the IEB is therefore difficult and can only be achieved using a special procedure (mapping). Unfortunately, the mapping does not allow a full transfer of the affected variables of coArb to VerBIS, which means that the attributes "n / a", "miscellaneous" or "missing" occur unusually frequently in some variables. In addition, frequency counts may show noticeable differences depending on whether the source of the data was coArb or VerBIS. Important limitations of the analysis potential are addressed in the respective variable description.

- 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.

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 under-recording have emerged for the authorised municipalities since 1 January 2005. Thus, data from the XASU source should only be analysed from 2007 onwards.

- 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', 'severe disability status', 'reason for 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 – activity performed' 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.

5.7 Participants-In-Measures History Files (MTH)

– The MTH is incomplete for measures with a start date before 01.01.2000.

– As of 01.01.2005 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 und 2.7).

– The MTH contains only notifications that are recorded in BA procedures. The use of these procedures in cooperation 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 reported via the XSocial standard, are not contained in the MTH or the NEPS-SC6-ADIAB.

– As a result of the reorganisation of the institutions responsible for implementing SGB-II, the documentation of participation in measures in the MTH may end or begin again when there is a change in the reporting procedure.[21]

– 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.

---

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

6 Description of variables

Frequencies and overviews of the individual values and labels of the variables in NEPS-SC6-ADIAB can be found under http://fdz.iab.de.

6.1 Identifiers

6.1.1 Person ID for the NEPS-SC6-ADIAB (ID_t)

Variable label	Person-ID for the NEPS-SC6-ADIAB
Variable name	ID_t
Category	identifiers
Origin	generated
Data type	Numerical
Hierarchy	none
Detailed description	The Person-ID indicates which observations belong to the same person. It also indicates which individuals correspond in the survey and the administrative data, which means that the ID_t is included in both the survey data and the administrative data. This identification variable is artificial, which is why it does not allow conclusions to be drawn regarding the characteristics of the person or original identifiers.

6.1.2 Establishment ID (betnr)

Variable label	Establishment ID
Variable name	betnr
Category	identifiers
Origin	BeH
Data type	numerical
Hierarchy	none
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 ff. and pp. 27-30.) 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: a) If the company has one office only, or if the company has one office only in one municipality, this office is the establishment and is given an establishment number.

	b) 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. c) 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 regard to the allocation of establishment numbers as part of the notification procedure for social security must be observed: a) 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. b) A **workplace** is a unit in which employees work and which is not allocated an establishment number according to the above-mentioned principles. c) A **company** as a term combines establishment premises and workplaces belonging to the same employer. d) 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. e) Establishment and establishment premises are synonyms; branch office is a synonym for subsidiary, district office, outsourced 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 characteristics (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 Observation counter per person (spell)

Variable label	counter per person
Variable name	spell
Category	generated technical variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH
Data type	numerical
Hierarchy	none

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.

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
Data type	numerical
Hierarchy	none
Detailed description	The variable indicates the data source.

6.2.3 Year (jahr)

Variable label	Year
Variable name	jahr
Category	generated technical variables
Origin	BeH
Data type	numerical
Hierarchy	none
Detailed description	This variable is only included in the 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 Establishment File.

6.3 Period of validity

6.3.1 Original start date of observation (begorig)

Variable label	Original start date
Variable name	begorig
Category	period of validity
Origin	BeH, LeH, LHG, ASU, XASU, MTH
Data type	date
Hierarchy	none
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 3.1.8.

<table>
<tr>
<td></td>
<td>

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.

</td>
</tr>
</table>

6.3.2 Original end date of observation (endorig)

Variable label	Original end date
Variable name	endorig
Category	period of validity
Origin	BeH, LeH, LHG, ASU, XASU, MTH
Data type	date
Hierarchy	none
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 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 3.1.8. **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). endorig indicates the end date of the new period.

6.3.3 Start date of split episode (begepi)

Variable label	Episode start date
Variable name	begepi
Category	generated period of validity
Origin	BeH, LeH, LHG, ASU, XASU, MTH
Data type	date
Hierarchy	none

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 End date of split episode (endepi)

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

6.4.3 Nationality (nation)

Variable label	Nationality
Variable name	nation
Category	personal variables

Origin	BeH, LeH, LHG, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
Detailed description	The variable contains the nation codes used by the Federal Statistical Office (Statistisches Bundesamt).
Notes	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.

6.4.4 Nationality, aggregated (nation_gr)

Variable label	Nationality, grouped
Variable name	nation_gr
Category	personal variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
Detailed description	The variable contains a grouped version of the nation codes used by the Federal Statistical Office.

6.4.5 Marital status (famst)

Variable label	Marital status
Variable name	famst
Category	personal variables
Origin	LeH, LHG, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
Detailed description	This variable describes the marital status. The variable in the LeH source has only two values (0 - not married, 1 - married), while in the LHG/ASU/XASU/MTH sources, a distinction is made between six values. The information from the sources was not compared.

6.4.6 Number of children (kind)

Variable label	Number of children
Variable name	kind
Category	personal variables
Origin	LeH, LHG, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
Detailed description	This variable has a different meaning depending on the data source. **1) LeH** The LeH source indicates the number of children aged under 16 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 |

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 restricts the quality of the data.

2) ASU, MTH

Here, the value of the variable corresponds to the number of children.

Until 30 June 2006, only up to nine children could be recorded. The value 0 does not exist. For observations prior to 30 June 2006, the zero value 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.

3) XASU, LHG

The variable shows the number of children aged under 16 in the benefit community (Bedarfsgemeinschaft). In the LHG sources, the value is valid for the entire original period.

6.4.7 Vocational training (ausbildung)

Variable label	Vocational training
Variable name	ausbildung
Category	personal variables
Origin	BeH, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
Detailed description	It must be taken into account that this variable has a different meaning depending on the data source: **1) BeH** For BeH observations the variable contains the vocational education and training reported by the employers in the employment notification procedure. The following values exist: 1 No vocational training 2 In-company voc. training/traineeship/external voc. training 11 Degree from a university of applied sciences 12 University degree In notifications using the new occupation code 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.

	"Changes in the vocational training status frequently occur at the same time as a change of establishment. This is because the notification data are compiled anew in the new firm. If, for example, an employee has gained a higher qualification via a part-time further training course while still working then this change of status is probably not recorded until he/she joins a new firm. It can generally be assumed that when a person is employed in a firm for a longer period, the personal data that they reported when they joined the firm is simply continued." (Meinken & Koch 2004, p. 63).
	A method for correcting missing values or temporal inconsistencies in the education and training data in the predecessor sample, the IABS, can be found in Fitzenberger et al. (2006) and in Drews (2006). Only the BeH data source is used for this, however, as this was the only data source with information on education and training in the IABS.
	2) ASU, XASU, MTH
	For these observations the vocational education or training completed most recently is reported. The following values exist:
	1 no completed vocational training 2 in-firm vocational training/external vocational training 3 full-time vocational school (Berufsfachschule) 4 technical college (Fachschule) 5 university of applied sciences (Fachhochschule) 6 university 7 vocational education/training not recognised in Germany 8 university degree not recognised in Germany
	Values 7 and 8 are only valid for the MTH data source.
Notes on quality	The number of missing values increases continuously over time. In the most recent data more than 40% of values are missing. Missing values occur especially frequently in the following person groups: persons in marginal part-time employment, persons working part-time, foreign employees and eastern German workers. The reason for this is that the variable is not of particular importance as regards social security (see Meinken & Koch, 2004, p. 63).
	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. The degree of completeness in the XASU data source is generally low.

6.4.8 School leaving qualification (schule)

Variable label	School leaving qualification
Variable name	schule
Category	personal variables

Origin	BeH, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
Detailed description	This variable contains the school leaving qualification. Different values are possible depending on the source. **1) BeH** The possible values change when the new occupation code is adopted (see Bertat et al., 2013). The values with the old occupation code are: 5 Grade-/lower secondary school with or without leaving certificate, intermediate school leaving certificate or equivalent qualification 8 Upper secondary school leaving certificate from a specialised upper secondary school (Fachoberschule), general upper secondary school leaving certificate, A-level equivalent, qualification for university 9 General upper secondary school leaving certificate, A-level equivalent, qualification for university With the new occupation code the values are: 1 No school leaving certificate 3 Lower secondary school certificate/grade school certificate 6 Intermediate school leaving certificate 8 Upper secondary school leaving certificate from a specialised upper secondary school/general upper secondary school leaving certificate, A-level equivalent, qualification for university **2) ASU, XASU, MTH** 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 Upper secondary school leaving certificate from a specialised upper secondary school (Fachoberschule) 9 General upper secondary school leaving certificate, A-level equivalent, qualification for university 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 has been decreasing continuously over time in the BeH and has stood at under 2/3 in recent years. In the XASU it has been increasing continuously and has stood at over 2/3 since 2012. In ASU and MTH the degree of completeness is good on the whole.

6.5 Information about employment, benefit receipt und job search

6.5.1 Daily wage, daily benefit rate (tentgelt)

Variable label	Daily wage/daily benefit
Variable name	tentgelt
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH
Data type	numerical
Hierarchy	none
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 information, however, may be considered less error-prone due to its insurance relevance.)The marginal part-time income threshold and the upper earnings limit for statutory pension insurance differ from year to year as well as between eastern and western Germany (the decisive factor is the location of the establishment). An overview of these limits and thresholds can be found under http://fdz.iab.de. A daily wage reported as 0 euros can be put down to "employment interruption notifications". During these periods, the employment relationship continues to exist in legal terms, but without pay. This is the case for periods of illness after the end of continued payment of wages, for periods of maternity leave and for sabbaticals.

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 start date prior to 1 January 1998 the daily benefit rate applies to working days, while for observations with an original start 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 Employment – 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
Hierarchy	none
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 30.11.2011 are reported using the new occupation code 2010 (KldB2010). These values are recoded to the KldB1988 by transferring the key area. 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.

| Note on quality | There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. |
| | In the XASU the occupation variable is not reported for almost the entire period available. |

6.5.3 Employment group – current/most recent (KldB 2010) (beruf2010_3)

Variable label	Occupational group - current/most recent (KldB 2010), 3-digit
Variable name	beruf2010_3
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
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 30.11.2011 are reported using the old occupation code 1988 (KldB 1988). 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
	The variable contains the occupation of the last job. See 1) with regard to the occupation code.
Note on quality	There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure.

6.5.4 Employment sub-group – current/most recent (KldB 2010) (beruf2010_4)

Variable label	Occupational sub-group - current/most recent (KldB 2010), 4-digit
Variable name	beruf2010_4
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, ASU, XASU, MTH
Data type	numerical
Hierarchy	none

Detailed description	**1) BeH**
	The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes apply for one employee, the employer is required to select the job title that best defines the main activity performed (see Bundesagentur für Arbeit, 2005, p. V).
	For this the employer encodes the employee's job in accordance with the "Classification of Occupations 2010" (Klassifikation der Berufe 2010, KldB2010, Bundesagentur für Arbeit, 2011). The occupational class consists of a 5-digit code and comprises about 1300 values. The less detailed occupational sub-group is recorded by the first four digits of the code. The skill level required for a job, which is recorded in the fifth digit of the codes in the KldB2010, is made available separately in the variable 'level of requirement' (niveau).
	Employment notifications with an end date earlier than 30.11.2011 are reported using the old occupation code 1988 (KldB 1988). 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
	The variable contains the occupation of the last job. See 1) with regard to the occupation code.
Note on quality	There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure.
Notes	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 occupational group (beruf2010_3) is the only occupation-related information using KldB2010 that is made available.

6.5.5 Level of requirement – current/most recent (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
Data type	numerical
Hierarchy	none
Detailed description	**1) BeH**
	The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes apply for one employee, the employer is required to select the job title that best defines the main activity performed (see Bundesagentur für Arbeit, 2005, p. V).
	For this the employer encodes the employee's job in accordance with the "Classification of Occupations 2010" (Klassifikation der Berufe 2010,

KldB2010, Bundesagentur für Arbeit, 2011). The occupational class consists of a 5-digit code and comprises about 1300 values. The less detailed occupational sub-group is recorded by the first four digits of the code. The skill level required for a job, which is recorded in the fifth digit of the codes in the KldB2010, is made available separately in the variable 'level of requirement' (niveau).

Employment notifications with an end date earlier than 30.11.2011 are reported using the old occupation code 1988 (KldB 1988). 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

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

Note on quality	There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure.

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
Hierarchy	none
Detailed description	The employee's occupational status during the notification period is reported by the employer in the "employment details". The variable "occupational status" distinguishes between full-time and part-time employees. The decisive factor here is the ratio between the contracted hours and the usual working hours in the establishment. For part-time employees the variable only records whether their working hours exceed a certain limit or not. Until 1978 this limit was 20 hours of work per week, between 1979 and 1987 it was 15 hours per week and since 1988 it has been 18 hours per week. The variable only provides actual details regarding the occupational status for full-time employees, distinguishing among other things between blue-collar and white-collar employees in full-time employment and trainees/apprentices. When the new occupation code was introduced, however, this distinction was no longer available. The variable 'teilzeit' therefore only distinguishes between full-time and part-time employment in the entire reporting period. No further information about the occupational status is used.
Note 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).

6.5.7 Employment status (erwstat)

Variable label	Employment status
Variable name	erwstat
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, LHG, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
Detailed description	This variable assumes 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 (cf. Deutsche BKK 2012, p. 31). The person group can be contained in employment notifications that refer to the years prior to 1999 but were not received until 1999 or later. For notifications which were received before 1999, an attempt is made to allocate the notifications to the person groups on the basis of certain rules and with the aid of the vocational education and training', 'occupational status and working hours' and 'occupation' variables as well as other information. In many cases, however, conclusive allocations are not possible. Since 1 April 1999, employees in marginal part-time employment have also been recorded in the DEÜV notification procedure. This person group can be distinguished via the values 109 and 209. For employees in marginal part-time employment, no data prior to the introduction of the notification obligation in 1999 could be collected. **2) LeH** For LeH observations, the variable 'employment status' contains the grouped benefit type. Thus, it is possible to distinguish whether a person receives unemployment benefit, unemployment assistance or maintenance allowance or whether contributions to private long-term care insurance are paid by the BA. **3) ASU/XASU**

	For ASU observations, the 'employment status' variable reports the job search status. Recipients of unemployment benefits (Unemployment Benefit I or II) over the age of 58 who receive benefits under the relaxed conditions according to Section 428 of Social Code Book III (or Section 65 Para. 4 of Social Code Book II) and individuals aged over 58 who are not benefit recipients and are not willing to be placed in employment in the sense of Section 252 Para. 8 Social Code Book VI are recorded as individuals seeking advice.
	The status 'without status' (statistics: 'not set') mainly implies individuals who cannot be expected to be activated or placed in employment in accordance with Section 10 SGB II. Individuals who are classed as unfit for work for more than 42 days but continue to receive Unemployment Benefit II are also recorded under this status in the system.
	In XASU observations, the variable 'employment status' has so far also contained the values 'not unemployed, but seeking work' as well as 'unemployed and simultaneously seeking work'.
	Since the notification logic might in many cases permit re-identification of the exact day of birth, the original information on the date was changed by applying the anonymisation procedure described in Section 3.1.8.
	4) LHG
	For LHG datasets, the 'employment status' variable shows whether the person is registered as an employable minor, an employable person of full age or not employable from the old-age pension threshold.
	Since the notification logic might in many cases permit re-identification of the exact day of birth, the original information on the date was changed by applying the anonymisation procedure described in Section 3.1.8.
	5) MTH
	For MTH observations the 'employment status' variable indicates the measure-type category. This is the highest level in the hierarchy of the measure-type classifications of the BA.
Note on quality	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 classed 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.

6.5.8 Transition zone (gleitz)

Variable label	Transition zone
Variable name	gleitz
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Hierarchy	none

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.

6.5.9 Temporary agency work (leih)

Variable label	Temporary agency work
Variable name	leih
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Hierarchy	none
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 30.11.2011.
Note on quality	There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure.

6.5.10 Fixed-term contract (befrist)

Variable label	Fixed-term contract
Variable name	befrist
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Hierarchy	none
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.
Note on quality	There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure.

6.5.11 Reason of cancellation/notification/termination (grund)

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

Hierarchy	none
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), whilst 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 LeH reports, the variable shows the reason for the termination of the unemployment benefits, unemployment assistance or maintenance. There is no information on the reasons for the start of the benefit receipt in the LeH, since the LeH is filled with the reports of the employment agencies to the health insurance concerning completed spells of benefit receipt. **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 community. If the Unemployment Benefit II receipt of a different member of the benefit community is discontinued, new observations for all members of the benefit community 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:

old → new	old → new	old → new	old → new
29 → 60	36 → 61	44 → 74	51 → 74
30 → 60	37 → 66	45 → 77	52 → 76

	31 → 61	38 → 66	46 → 67	53 → 68
	32 → 60	39 → 71	47 → 67	54 → 78
	33 → 60	40 → 69	48 → 78	
	34 → 60	42 → 65	49 → 69	
	35 → 60	43 → 70	50 → 75	

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.

Note on quality	The proportion of valid values (degree of completeness) for the reason for notification in the LHG data sources is very small (< 20%) across all years. The degree of completeness is also low in the XASU, but improves over time.

6.5.12 Employment status prior to job search (estatvor)

Variable label	Employment status prior to job search
Variable name	estatvor
Category	information on employment, benefit receipt and job search
Origin	ASU, XASU
Data type	numerical
Hierarchy	none
Detailed description	For ASU and XASU observations, this variable shows the employment status prior to the job search activity. From 12/2002 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
Note on quality	The proportion of valid values (degree of completeness) increases for XASU observations from approx. 8 % to approx. 78 % between 2005 and 2013.

6.5.13 Employment status after job search (estatnach)

Variable label	Employment status after job search
Variable name	estatnach
Category	information on employment, benefit receipt and job search
Origin	ASU
Data type	numerical
Hierarchy	none

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

6.5.14 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
Hierarchy	none
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.
Note 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.

6.5.15 Type of termination of last job (art_kuend)

Variable label	Reason for end of previous employment
Variable name	art_kuend

Category	information on employment, benefit receipt and job search
Origin	ASU, MHT
Data type	numerical
Hierarchy	none
Detailed description	This variable describes how the last employment or training relationship was terminated before a new period of job search. It can therefore be used to identify job-to-job placements.

6.5.16 Desired working hours of the job sought (arbzeit)

Variable label	Desired hours of the job sought
Variable name	arbzeit
Category	information on employment, benefit receipt and job search
Origin	ASU, MTH
Data type	numerical
Hierarchy	none
Detailed description	During the placement procedure, jobseekers indicate how many working hours the job they are seeking should have.

6.5.17 Residual claim / planned duration (restanspruch)

Variable label	Residual claim/planned duration
Variable name	restanspruch
Category	information on employment, benefit receipt and job search
Origin	LeH, MTH
Data type	numerical
Hierarchy	none
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.

6.5.18 Type of institution (traeger)

Variable label	Type of institution
Variable name	traeger
Category	information on employment, benefit receipt and job search
Origin	LHG, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
Detailed description	The variable contains the type of institution responsible for implementing Unemployment Benefit II (LHG), the type of institution providing the measure (MTH) 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 institution responsible for Unemployment Benefit 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.

6.5.19 Start date of unemployment (alo_beg)

Variable label	Start date of unemployment
Variable name	alo_beg
Category	information on employment, benefit receipt and job search
Origin	LeH, LHG, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
Detailed description	The variable reports the start date of an uninterrupted sequence of unemployment periods and is valid at the beginning of the observation. The following gaps do not result in an interruption of the unemployment period: • any gap lasting seven days or less • periods of illness lasting up to 42 days (ASU) No information about illnesses is contained in XASU observations, which is why it cannot be taken into account in the calculations.

6.5.20 Duration of unemployment (alo_dau)

Variable label	Duration of unemployment
Variable name	alo_dau

Category	information on employment, benefit receipt and job search
Origin	LeH, LHG, ASU, XASU, MTH
Data type	numerical
Hierarchy	none
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 (ASU) When calculating the duration these gaps are not added, however. No information about illnesses is contained in XASU observations, which is why it cannot be taken into account in the calculations. 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
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 (Kreis-schlüssel) show the code for the federal state (Bundesland), positions 1-3 indicate the regional authority (Regierungsbezirk), and positions 1-5 show the district authority (Kreis). Federal states without a regional authority have a 0 in the third position. In the BeH, the place of residence is determined at the end of each year and added consistently to all datasets of a year. For the LHG, XLHG and XASU sources, the place of residence applies to the period of the original observation. For the ASU and LeH, the variable contains the place of residence at the beginning of the original period of time.

	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 2014 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 2014. 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	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 (wo_bula) is shown as regional information.

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
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 observations, the federal state of the place of residence is only available for the years from 1999 onwards. Further information on the district of the place of residence can be found under Place of residence: district (Kreis) (wo_kreis).

6.6.3 Place of residence – employment agency (wo_aa)

Variable label	Place of residence: employment agency (Arbeitsagentur)
Variable name	wo_aa
Category	location data
Origin	BeH, LeH, LHG, XLHG, ASU, XASU
Data type	numerical
Hierarchy	regional directorate | employment agency
Detailed description	This variable contains the employment agency of the employee's/BA client's place of residence from 1999 onwards. 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 and MTH, the variable contains the place of residence at the start of the period of unemployment or job search. 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 employment agency was recoded to the territorial allocation of 31 December 2014 for all data sources, i.e. in all calendar years, a place of residence is assigned to an employment agency area in accordance with the boundaries that the employment agency area had on 31 December 2014. As the boundaries of the employment agency areas 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. Prior to 1999 the variable contains the employment agency that last processed the case of benefit receipt. It is not determined from the residence address and is only available for the LeH data source. Moreover, it is not possible to update the territorial allocations. This must be taken into account when conducting calculations over time (e.g. regional unemployment figures).
Note on quality	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 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 (wo_rd)

Variable label	Place of residence: regional directorate (Regionaldirektion)
Variable name	wo_rd
Category	location data
Origin	BeH, LeH, LHG, ASU, XASU, MTH
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 Economic activity 73, 3-digit code (w73_3)

Variable label	classification of economic activities 73
Variable name	w73_3
Category	establishment variables
Origin	BeH
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 Economic activity 93, 5-digit code (w93_5)

Variable label	classification of economic activities 93, sub-classes
Variable name	w93_5
Category	establishment variables
Origin	BeH
Data type	numerical
Hierarchy	section (1-digit code)	division (2-digit code)	group (3-digit code)	class (4-digit code)	sub-class (5 digit code) of economic activity
Detailed description	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.
Note on quality	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 Economic activity 93, 3-digit code (w93_3)

Variable label	classification of economic activities 93, groups
Variable name	w93_3
Category	establishment variables
Origin	BeH

Data type	numerical
Hierarchy	section (1-digit code) | division (2-digit code) | group (3-digit code) | class (4-digit code) | sub-class (5 digit code) of economic activity
Detailed description	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 Economic activity 03, 5-digit code (w03_5)

Variable label	classification of economic activities 03, sub-classes
Variable name	w03_5
Category	establishment variables
Origin	BeH
Data type	numerical
Hierarchy	section (1-digit code) | division (2-digit code) | group (3-digit code) | class (4-digit code) | sub-class (5-digit code) of economic activity
Detailed description	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.
Note on quality	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 Economic activity 03, 3-digit code (w03_3)

Variable label	classification of economic activities 03, groups
Variable name	w03_3
Category	establishment variables

Origin	BeH
Data type	numerical
Hierarchy	section (1-digit code) | division (2-digit code) | group (3-digit code) | class (4-digit code) | sub-class (5 digit code) of economic activity
Detailed description	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 Economic activity 08. 5-digit code (w08_5)

Variable label	classification of economic activities 08, sub-classes
Variable name	w08_5
Category	establishment variables
Origin	BeH
Data type	numerical
Hierarchy	section (1-digit code) | division (2-digit code) | group (3-digit code) | class (4-digit code) | sub-class (5-digit code) of economic activity
Detailed description	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.
Notes	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 Economic activity 08, 3-digit code (w08_3)

Variable label	classification of economic activities 08, groups
Variable name	w08_3
Category	establishment variables
Origin	BeH
Data type	numerical

Hierarchy	section (1-digit code)	division (2-digit code)	group (3-digit code)	class (4-digit code)	sub-class (5 digit code) of economic activity
Detailed description	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 Economic activity 73 generated – completed by extrapolation / imputation (w73_3_gen)

Variable label	w73_3 completed by extrapolation/imputation
Variable name	w73_3_gen
Category	establishment variables
Origin	BeH
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 Economic activity 73 generated – type of completion (group_w73_3)

Variable label	Type of imputation w73_3
Variable name	group_w73_3
Category	establishment variables
Origin	BeH
Data type	numerical
Hierarchy	None
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 Economic activity 93 generated – completed by extrapolation / imputation (w93_3_gen)

Variable label	w93_3 completed by extrapolation/imputation
Variable name	W93_3_gen
Category	establishment variables
Origin	BeH
Data type	numerical
Hierarchy	section (1-digit code) | division (2-digit code) | group (3-digit code) | class (4-digit code) | sub-class (5 digit code) of economic activity
Detailed description	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 Kind of completion w93_3 (group_w93_3)

Variable label	Type of imputation w93_3
Variable name	group_w93_3
Category	establishment variables
Origin	BeH
Data type	numerical
Hierarchy	None
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 Year of first appearance of establishment number (grd_jahr)

Variable label	year of first appearance
Variable name	grd_jahr
Category	establishment variables
Origin	BeH
Data type	numerical
Hierarchy	none

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.

6.7.13 First appearance of establishment number (grd_dat)

Variable label	first appearance
Variable name	grd_dat
Category	establishment variables
Origin	BeH
Data type	date
Hierarchy	none
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.
Notes	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).

6.7.14 Year of last appearance of establishment number (lzt_jahr)

Variable label	year of last appearance
Variable name	lzt_jahr
Category	establishment variables
Origin	BeH
Data type	numerical
Hierarchy	none

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).

6.7.15 Last appearance of establishment number (lzt_dat)

Variable label	last appearance
Variable name	lzt_dat
Category	establishment variables
Origin	BeH
Data type	date
Hierarchy	none
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 reasons for this are an "arbitrary change of the establishment number following a change of ownership or a change in the legal form of the establishment", the "outsourcing of parts of the firm under a new establishment number" or other administrative changes (see Bender et al. 1996 or Bundesagentur für Arbeit 2007, pp. 9-11).
Notes	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).

6.7.16 Total number of employees (az_ges)

Variable label	no. employees
Variable name	az_ges
Category	generated establishment variables
Origin	BeH
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. Part-time employees and dormant employment relationships (daily wage of zero) are also included.

6.7.17 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	BeH
Hierarchy	none
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, 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.

6.7.18 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	BeH
Hierarchy	none
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.

6.7.19 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	BeH
Data type	numerical
Hierarchy	none
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 8.2 in Schmucker et al. 2016). These data were then aggregated at establishment level.

6.7.20 Place of work: district (ao_kreis)

Variable label	Place of work: district (Kreis)
Variable name	ao_kreis
Category	location data
Origin	BeH
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), positions 1-3 indicate the regional authority (Regierungsbezirk), and positions 1-5 show the district authority (Kreis). Federal states without a regional authority have a 0 in the third position.

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 2014, 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 2014. 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. |
| Notes | 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. |

6.7.21 Place of work: federal state (Bundesland) (ao_bula)

Variable label	Place of work: federal state (Bundesland)
Variable name	ao_bula
Category	location data
Origin	BeH
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.

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
Hierarchy	none
Detailed description	This variable is only included in the matching file. For every respondent in the NEPS-SC6 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
Hierarchy	none
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
Hierarchy	none
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.

7 References

Antoni, M., Ganzer, A., Berge Philipp v. (2016). Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975-2014." FDZ-Datenreport 04/2016 (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.

Antoni, M., & Seth, S. (2012). ALWA-ADIAB–Linked individual survey and administrative data for substantive and methodological research. Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 132(1), 141-146.

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 (Access: 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.

Deutsche BKK. (2016). Ratgeber Sozialversicherung 2016, Wolfsburg. URL: http://docplayer.org/14261631-Ratgeber-sozialversicherung-2016.html (Zugriff: 05.04.2018)

Drews, N. (2006). Qualitätsverbesserung der Bildungsvariable in der IAB-Beschäftigten-stichprobe 1975-2001. FDZ Methodenreport 05/2006 (de), Nürnberg.

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. (2005). Imputation rules to improve the education variable in the IAB employment subsample. Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 126(3), 405-436.

Hammon, A., Zinn, S., Aßmann, C., & Würbach, A. (2016). Samples, Weights, and Nonresponse: The Adult Cohort of the National Educational Panel Study (Wave 2 to 6). NEPS Survey Papers.

Kleinert, C., Matthes, B., Antoni, M., Drasch, K., Ruland, M., & Trahms, A. (2011). ALWA–new life course data for Germany. Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, 131(4), 625-634.

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., Seth, S., Ludsteck, J., Eberle, J., & Ganzer, A. (2016). Betriebs-Historik-Panel 1975-2014. FDZ-Datenreport 03/2016 (de), Nürnberg.

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

Statistik der Bundesagentur für Arbeit. (2013). Neue Erhebungsinhalte „Arbeitszeit", „ausgeübte Tätig-keit" sowie „Schul- und Berufsabschluss" in der Beschäftigungsstatistik. Methodenbericht.

Statistisches Bundesamt. (2002). Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ93).

Statistisches Bundesamt. (2003). Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003).

Statistisches Bundesamt. (2008). Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

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.

8 Appendix

8.1 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.

9   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
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)
DEUV	Verordnung über die Erfassung und Übermittlung von Daten für die Träger der Sozialversicherung – Daten-erfassungs- und –übermittlungs-verordnung	Data Collection and Transmission Regulation - regulation on the collection and transmission of data for the social security agencies
DEVO	Zweite VO über die Erfassung von Daten für die Träger der Sozialversicherung und für die BA – Datenerfassungs-Verordnung –	Data Collection Regulation - second regulation on the collection of data for the social security agencies and for the Federal Employment Agency
DUVO	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
RDC	Forschungsdatenzentrum	Research Data Centre
FELEG	Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbs-tätigkeit	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
IAB	Institut für Arbeitsmarkt- und Berufsforschung	Institute for Employment Research
IAB-RDC	Forschungsdatenzentrum der Bundesagentur für Arbeit am IAB	Research Data Centre of the Federal Employment Agency at the IAB
IABS	IAB-Beschäftigtenstichprobe	IAB Employment Samples
IEB	Integrierte Erwerbsbiographien	Integrated Employment Biographies

ISIC	International Standard Industrial Classification of All Economic Activities	International Standard Industrial Classification of All Economic Activities
LeH	Leistungsempfänger-Historik	Benefit Recipient History
LHG	Leistungs-Historik Grundsicherung	Unemployment Benefit II Recipient History
LIfBi	Leibniz Institut für Bildungsverläufe	Leibniz Institute for Educational Trajectories
LIfBi-RDC	Forschungsdatenzentrum des Leibniz Institut für Bildungsverläufe	Research Data Center of the Leibniz Institute for Educational Trajectories
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	National Educational Panel Study	National Educational Panel Study
NEPS-SC6	NEPS Startkohorte 6	NEPS Starting Cohort 6
SchwbG	Gesetz zur Sicherung der Eingliederung Schwerbehinderter in Arbeit, Beruf und Gesellschaft – Schwerbehindertengesetz-	law to guarantee the integration of persons with severe disabilities into employment and society – Severely Disabled Persons Act
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
zkT	Zugelassener kommunaler Träger	Authorised municipalities

Corresponding author:
Nadine Bachbauer
Institute for Employment Research (IAB)
Research Data Centre (FDZ)
Regensburger Str. 104
D-90478 Nürnberg
Phone: +49-911-179-9579
Email: Nadine.Bachbauer2@iab.de