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

FDZ-DATENREPORT

Documentation of labour market data

05|2025 EN Children of Immigrants Longitudinal Survey in Four European Countries - Germany linked to administrative data of the IAB (CILS4EU-DE-ADIAB) 1975-2023

Manfred Antoni (IAB), Jörg Dollmann (MZES), Alexandra Schmucker (IAB), Markus Weißmann (MZES)

Bundesagentur für Arbeit

Children of Immigrants Longitudinal Survey in Four European Countries - Germany linked to administrative data of the IAB (CILS4EU-DE-ADIAB) 1975-2023

Manfred Antoni (IAB), Jörg Dollmann (MZES), Alexandra Schmucker (IAB), Markus Weißmann (MZES)

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

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

Contents

Zusammenfassung

Dieser Datenreport beschreibt die verknüpften Befragungsdaten des Children of Immigrants Longitudinal Survey in Four European Countries - Germany mit administrativen Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) 1975-2023.

Abstract

This data report describes the linked data of the Children of Immigrants Longitudinal Survey in Four European Countries - Germany Sample with administrative data of the Institute for Employment Research (IAB).

Keywords

German administrative micro data, labour market data, data manual, Children of Immigrants Longitudinal Survey in Four European Countries, CILS4EU-DE, survey data, record linkage, integration of immigrants and their children

Acknowledgements

We would like to thank our colleagues in the Research Data Centre (FDZ) as well as the department Data and IT-Management (DIM) of the Institute for Employment Research for their cooperation and support. Individual passages of this Datenreport were adopted from IAB-internal data documentations by DIM.

Data availability

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

1 Introduction and outline

1.1 Introduction

This data report describes the data product Children of Immigrants Longitudinal Survey in Four European Countries -– Germany" linked to administrative data of the IAB 1975-2023 (CILS4EU-DE-ADIAB 7523), which is provided cooperatively by the Mannheim Centre for European Social Research (MZES) of the University of Mannheim and the Institute for Employment Research (IAB). CILS4EU-DE-ADIAB consists of the survey data of respondents of the „Children of Immigrants Longitudinal Survey in Four European Countries – Germany" (CILS4EU-DE) and administrative data of the IAB, in case the respondents consented to their records being linked and could be identified in the IAB data.

CILS4EU-DE is the German extension of the international project CILS4EU, which was funded by the NORFACE ERA NET Plus Migration in Europe programme and aimed to analyse the complex processes of integration among immigrant youth and youth with a migration background. Since 2014, CILS4EU-DE has been funded through the German Research Foundation's (DFG) long-term programme.

CILS4EU conducted three waves of surveys among adolescents in England, Germany, the Netherlands, and Sweden from 2010 to 2013. Respondents were first surveyed in 2010, when they were 14 or 15 years old, and then again in the following two years, during their transition from adolescence to adulthood. The German part of CILS4EU was continued as CILS4EU-DE since wave 4 (2014). At the core of the study is the analysis of complex processes of the integration of immigrant children and children with a migration background. The survey covers various indicators of structural, social, cultural, and emotional integration—such as educational and labour market outcomes, friendships and romantic relationships, language proficiency, attitudes, and national or ethnic identification—providing a rich basis for an in-depth analysis of their complex causal relations. In addition, one parent was surveyed at the beginning of the study, adding an intergenerational perspective and allowing for the examination of integration and transmission processes across generations. Moreover, the sample makes it possible to compare integration processes between ethnic groups.

The first wave of the German part of the study was conducted in the fall of 2010 among students attending Grade 9 in general education schools. The initial sample was drawn using a three-stage sampling design (see CILS4EU, 2016). In the first stage, all general schools in Germany were sampled. To ensure a sufficiently high number of students with a migration background, schools with a high share of immigrant students were oversampled. In total, 144 schools were selected. In the second stage, two classes in the ninth grade in the selected schools were drawn at random. In the third stage, all students in these classes were invited to participate in the survey. A small number of students was excluded because of mental or physical limitations or because an interview was not possible or difficult due to language problems. The first and second waves of the survey (2010 and 2011) were administered in the participating schools. The parent interviews in Wave 1 were conducted exclusively using postal questionnaires or telephone interviews. In Wave

2, students who were absent, had changed schools or classes, or had left the school were followed up via telephone, postal, or online questionnaires. From Wave 3 onwards, all survey waves (except for Wave 6 in 2016; see next paragraph) were administered using one of these three modes.

In addition to the sample from 2010, a comparable refreshment sample was drawn in 2016 to account for panel attrition (see Schiel et al., 2016). The target population comprised all individuals residing in Germany who were born between 1994 and 1996. In the first sampling stage, all municipalities in Germany served as the sampling frame, from which 62 two municipalities were randomly selected. The second-stage sampling unit comprised all residents born between 1994 and 1996. To arrive at a distribution by migration background comparable to the sample in Wave 1, residents with a migration background were oversampled. We determined whether residents had a migration background using a linguistic analysis of the first and last names (onomastic method) of the individuals from the relevant sampling units, which were retrieved from the lists of the municipal registration offices. The selected individuals were invited to participate in the survey and, upon consent, were interviewed at home via computer-assisted interviews. In Wave 6, the initial sample was interviewed in the same manner.

The linked data product contains the reduced version (download version) of the CILS4EU and CILS4EU-DE datasets.

The administrative data originate from the Integrated Employment Biographies (IEB) of the IAB. In the present case, they contain data on all persons interviewed for CILS4EU-DE, who agreed to be linked and could be linked, if they showed one of the following statuses at least once during the observation period:

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

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

While the Employee History (Beschäftigtenhistorik - BeH) is the origin of the information on employment subject to social security and marginal part-time employment, the receipt of benefits in accordance with Social Code Book III (SGB III) and Social Code Book II (SGB II) is recorded in the Benefit Recipient History (Leistungsempfängerhistorik - LeH) and the Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung - LHG). The Jobseeker Histories (Arbeitsuchendenhistoriken - ASU and XASU) are the data source for the periods of job search recorded by the BA or by municipal institutions responsible for implementing SGB II, while participation in employment and training measures is recorded in the Participants-in-Measures History Files (Maßnahmeteilnahmehistoriken – MTH and XMTH).

In addition to this individual data, information on the employing establishments can also be linked from the Establishment History Panel (BHP) of the IAB (Ganzer et al., 2024).

This Datenreport is structured as follows. In addition to the introduction, Chapter 1 contains information on data access as well as an outline of the data, the volume structure and a list of variables. A description of the individual data sources can be found in Chapter 2. Data preparation and sampling are discussed in Chapters 3. Chapters 4 and 6 contain a description of data linkage and data quality, while the individual variables are described in Chapter 6.

1.2 Data use

1.2.1 Data access

The CILS4EU-DE-ADIAB data may only be analysed in the context of a research visit at one of the locations of the FDZ and subsequent remote data execution.

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

Once approval has been granted, a data use agreement for the reduced versions of CILS4EU (waves 1-3: https://search.gesis.org/research_data/ZA5656) and CILS4EU-DE (waves 4-9: https://search.gesis.org/research_data/ZA6656) data must be concluded with the GESIS Leibniz Institute for Social Sciences in Mannheim (https://search.gesis.org/research_data/ZA6655), if this has not already been done. As soon as a data use agreement with GESIS has been concluded for the requested project, a data use agreement will be concluded between the researcher's institution and the IAB. Details on how to apply for the linked data set and the data processing options can be found on the IAB-FDZ website. Data access can therefore only begin after agreements have been concluded with both institutions for the requested project.

To answer questions about CILS4EU-DE-ADIAB, users can contact the staff of the RDCs at IAB and the University of Mannheim. The division of responsibilities between the two institutions includes that questions on CILS4EU-DE data should exclusively be sent to the University of Mannheim, and questions on the administrative data, technical aspects of data linkage or access to the linked data exclusively to the IAB-RDC.

1.2.2 Data management

The survey data of CILS4EU-DE and the administrative data of the IAB are stored in separate files, which makes the structure of the data clear and saves storage capacity. The personal data from both data sources can be merged via the person identification variable (youthid), which is also included in the administrative person data.

A further element of CILS4EU-DE -ADIAB is the linkage file, which contains the individual ID as well as technical characteristics with information on whether the linkage with the administrative data was successful and which linkage method was used. The file also contains the respondents of

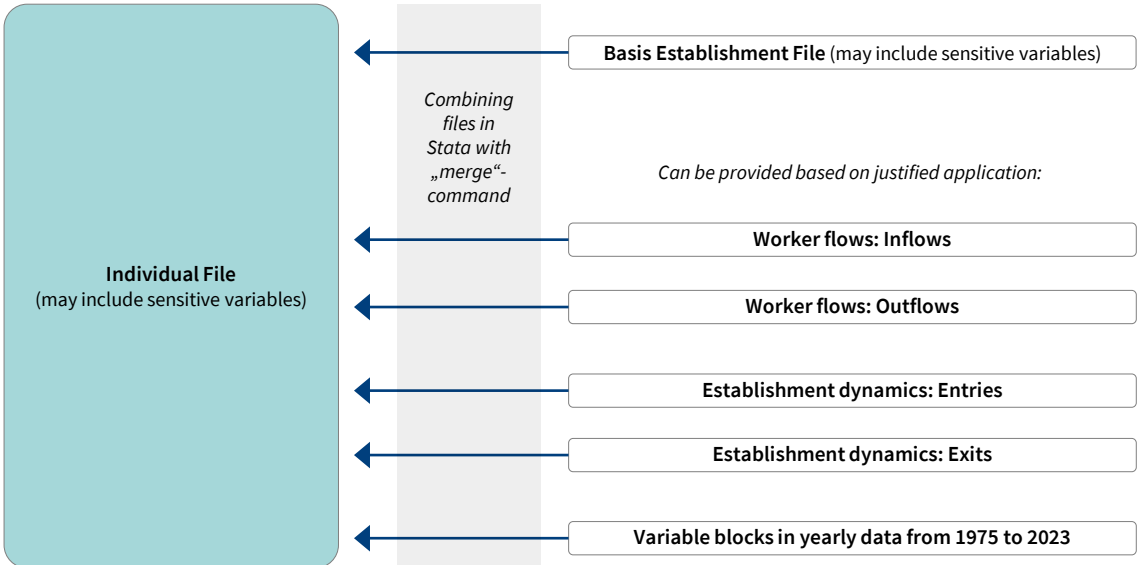
CILS4EU-DE for whom no administrative data exists or which did not consent to data linkage. Thus, analyses on non-consent and non-linkage can be carried out on the basis of all participants of CILS4EU-DE.

The administrative data, which include both German and English labels,[1] have a modular structure and are stored in several files. One module, which is henceforth called the Individual File, contains identifiers (individual IDs and establishment IDs), the personal variables, information on employment, benefit receipt and job search activity, variables regarding place of residence, and technical variables. These personal data includes daily data for the period 1975 to 2020.

A second module, the Basis Establishment File, contains the establishment number, the year, and variables regarding the place of work and economic activities in aggregate form as well as other establishment information as of the reference date of 30 June. The datasets for eastern Germany are available from 1992 onwards.

Establishment variables and individual variables are therefore organised separately, which makes the structure of the data clear and saves storage space (see Figure 1). The Basis Establishment File and the Individual File are linked via the establishment number and the year of the dataset. Please note that the variable, which is necessary for linking the two modules, “establishment number”, is available in both the Individual File and the Basis Establishment File.

Figure 1: Storage of the administrative data



1.2.3 Sensitive variables and additional variables

Certain, so-called sensitive variables, which could facilitate the de-anonymisation of individuals or establishments, are only disclosed in the original if this is necessary for the analysis objective and is explicitly justified in the application for data access. It should be noted that the sensitive variables are usually already included in the data in a coarsened version. The sensitive variables are only made available in detail if the information in the coarsened state is insufficient to achieve the research objective. When a dataset is provided with sensitive variables, the corresponding non-

[1] With the Stata command *label language en* / *label language de* labels can be switched to English or German.

sensitive variables are deleted if they can be generated by the users themselves, e.g. federal state from district information.

The variables which are particularly sensitive from the viewpoint of data protection legislation are:

Administrative Individual File:

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

Basis Establishment File:

- place of work: district (Kreis) (ao_kreis)
- economic activity 08 - sub-class of economic activity (five-digit code) (w08_5)

Upon reasoned request, further establishment variable blocks from the Establishment History Panel (Betriebs-Historik-Panel - BHP) (see Figure 1) can be provided. In addition, the extension files Worker flows (inflows/outflows) and Establishment dynamics (entries/exits) can be applied for with a separate justification.

These additional variable blocks or extension files are:

Variable blocks from the BHP (annual waves):

- General employment structure (az_f az_reg az_azubi az_atz az_tz az_f_vz az_f_tz az_reg_vz)
- Structure of employees by educational and vocational qualifications (az_gq az_mq az_hq az_gq_vz az_mq_vz az_hq_vz)
- Employee age structure (az_15_19 az_15_19_vz az_20_24 az_20_24_vz az_25_29 az_25_29_vz az_30_34 az_30_34_vz az_35_39 az_35_39_vz az_40_44 az_40_44_vz az_45_49 az_45_49_vz az_50_54 az_50_54_vz az_55_59 az_55_59_vz az_60_64 az_60_64_vz az_ab65 az_ab65_vz alter_mw alter_mw_vz)
- Research and development activities (az_ingnat)
- Number of employees with non-standard job types (az_leih az_bfr)
- Structure of employees by nationality (az_d az_d_vz az_eu)
- Structure of employees by Blossfeld occupational group (az_bf_agr az_bf_emb az_bf_edi az_bf_evb az_bf_qmb az_bf_qdi az_bf_qvb az_bf_tec az_bf_semi az_bf_ing az_bf_prof az_bf_man)
- Structure of employees by level of requirement (az_niv1 az_niv2 az_niv3 az_niv4)
- Wage structure of full-time employees (az_zens te_imp_med te_imp_p25 te_imp_p75 te_imp_mw_f te_imp_med_f te_imp_med_m te_imp_mw_gq te_imp_med_gq te_imp_mw_mq te_imp_med_mq te_imp_mw_hq te_imp_med_hq te_imp_med_uq te_imp_mw_d te_imp_med_d te_imp_med_a te_imp_sum)

Extension files of the BHP:

- Extension file – Worker flows (ein_ges ein_f ein_reg ein_gf ein_azubi ein_vz ein_tz ein_f_vz ein_f_tz ein_reg_vz ein_wdr ein_wdr_f ein_bw ein_bw_f ein_bf_agr ein_bf_emb ein_bf_edi ein_bf_evb ein_bf_qmb ein_bf_qdi ein_bf_qvb ein_bf_tec ein_bf_semi ein_bf_ing ein_bf_prof ein_bf_man ein_15_19 ein_20_24 ein_25_29 ein_30_34 ein_35_39 ein_40_44 ein_45_49 ein_50_54 ein_55_59 ein_60_64 ein_ab65 aus_ges aus_f aus_reg aus_gf aus_azubi aus_vz aus_tz aus_f_vz aus_f_tz aus_reg_vz aus_temp aus_bw aus_bw_f aus_bf_agr aus_bf_emb aus_bf_edi aus_bf_evb aus_bf_qmb aus_bf_qdi aus_bf_qvb aus_bf_tec aus_bf_semi aus_bf_ing aus_bf_prof aus_bf_man aus_15_19 aus_20_24 aus_25_29 aus_30_34 aus_35_39 aus_40_44 aus_45_49 aus_50_54 aus_55_59 aus_60_64 aus_ab65 aus_senio_1 aus_senio_2 aus_senio_3)
- Extension file – Entry and exit (eintritt besch besch_vor status_vor inflow austritt besch besch_nach status_nach outflow)

More information on the BHP can be found at https://fdz.iab.de/en/betriebsdaten/establishment-history-panel-bhp-version-7523-v1/ and in the FDZ data report 09/2024 (Ganzer et al., 2024).

In addition, person- and establishment-specific effects (AKM effects) are available for the CILS4EU-DE-ADIAB 7523 for the years 2001 to 2023. The variables feff_1985_1992 and feff_1993_2000 as well as peff_1985_1992 and peff_1993_2000 are not provided as CILS4EU-DE-ADIAB only contains information on employment relationship and employers from 2007 onwards. Further information can be found in the FDZ Methodenreport 03/2025 (Lochner/Wolter, 2025). For new projects the AKM effects can be requested in the field for sensitive variables and extension files. For running projects, the AKM effects can be applied for by e-mail to iab.fdz@iab.de.

1.2.4 Filenames of the administrative Data and Linkage File

Individual file

CILS4EU-DE-ADIAB_7523_v1.dta

Linkage file

CILS4EU-DE-ADIAB_7523_v1_key.dta

Basis Establishment File

CILS4EU-DE ADIAB_7523_v1_bhp_basis_v1.dta

Extension files BHP

Core datasets with blocks of variables

CILS4EU-DE-ADIAB_7523_v1_bhp_v1_yyyy.dta, yyyy = 2007 - 2023

Worker flows

CILS4EU-DE-ADIAB_7523_v1_bhp_inflow_v1.dta

CILS4EU-DE -ADIAB_7523_v1_bhp_outflow_v1.dta

Establishment dynamics

CILS4EU-DE-ADIAB_7523_v1_bhp_entry_v1.dta

CILS4EU-DE -ADIAB_7523_v1_bhp_exit_v1.dta

Person- and establishment-specific effects (AKM effects)

CILS4EU-DE-ADIAB_7523_v1_akm_estab_8523_v1.dta

CILS4EU-DE -ADIAB_7523_v1_akm_pers_8523_v1.dta

1.3 Changes as compared to previous version

Not applicable

1.4 Outline

Table 1: Outline

Category	Description
Topics/ groups of variables	**Survey data** Socio-demographic information Information on the socio-economic background of the family; household composition; living situation; retrospective residence history; information on origin and migration history of respondents, their parents and grandparents; citizenship. Structural integration Current situation (school, vocational training, employment); school situation; attitudes towards school; school performance; educational and occupational aspirations. Cultural integration Achievement tests, self-assessed language competencies and use in German and a further language; tolerance towards different ways of life; sexuality norms; leisure time activities; knowledge about and interest in politics of country of origin. Emotional integration Identification with Germany and country of origin; religion, religiosity and religious behaviour; perceived discrimination; attitudes towards different groups of origin; attitudes towards integrations; political participation; trust. Well-being Life satisfaction; subjective well-being; health-related behaviour; health problems; body height and body weight. **Employee History (Beschäftigtenhistorik - BeH):** Annual notifications and end-of-employment notifications submitted to the social security agencies for employees covered by social security and employees in marginal part-time employment. **Benefit Recipient History (Leistungsempfängerhistorik - LeH):** Information on benefit receipt in accordance with Social Code Book III (SGB III) for recipients of unemployment benefit, unemployment assistance and maintenance allowance.

Category	Description
	Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung - LHG): Data on individuals in receipt of basic social security benefits in accordance with Social Code Book II (SGB II) (Types of institution: cooperation of employment agencies and municipalities/joint facilities, separated responsibilities/municipalities exercising their duties separately, authorised municipalities) **Jobseeker History (Arbeitsuchendenhistorik - ASU):** Information on job search activities that are recorded in BA procedures **Jobseeker History from XSozial-BA-SGB II (Arbeitsuchendenhistorik aus XSozial-BA-SGB II - XASU):** Information on job search activity reported to the BA by authorised municipalities via the transmission standard XSozial-BA-SGB II **Participants-in-Measures History File (Maßnahmeteilnahmehistorik - MTH):** Information on participation in employment and training measures (not including measures of authorised municipalities) **Participants-in-Measures History File from XSozial-BA-SGB II (Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II - XMTH):** Information on participation in employment and training measures reported to the BA by authorised municipalities via the transmission standard XSozial-BA-SGB II
Data units	Survey data: First sample: Students who attended Grade 9 of a general education school in Germany in the fall of 2010. Oversampling of schools with a high share of immigrant students. Second sample: Individuals born between 1994 and 1996 who were registered in Germany in 2016. Oversampling of individuals with migration background (determined via linguistic analysis of first and last name; onomastic method) Administrative personal data: Employees covered by social security (including marginal part-time employees from 1999 onwards), benefit recipients, jobseekers, participants in measures, establishments
Number of cases	3,594 individuals 72,832 original observations 109,550 non-overlapping observations (after episode splitting)
Period covered	Survey data: wave 1: 2010/2011 wave 2: 2011/2012 wave 3: 2012/2013 wave 4: 2014 wave 5: 2015 wave 6: 2016/2017 wave 7: 2018 wave COVID: 2020 wave 8: 2020 wave 9: 2022 Administrative data:

Category	Description
	The period covered depends on the data source. BeH: 01.01.1975 - 31.12.2023 (until 2020: 36-months-file 2021 and 2022: 18-months-file, 2023: 6- months-file) LeH: 01.01.1975 - 31.12.2023 ASU: 01.01.1997 - 31.12.2023 LHG: 01.01.2005 - 31.12.2023 XASU: 01.01.2005 - 31.12.2023 MTH: 01.01.2000 - 31.12.2023 XMTH: 01.01.2005 - 31.12.2023
Time reference	Survey data: Time of survey; since last survey; spell data retrospectively Administrative personal data Exact to the day Administrative establishment data Reporting date 30 June
Regional structure	Survey data: Germany Administrative data: German federal states (Bundesländer), districts (Kreise)
Date of territorial allocation	Administrative data: Territorial allocation updated as of 31.12.2023
Data collection method	Survey data: Wave 1: Survey of students PAPI (in-school), parents PAPI or CATI (in-home) Wave 2: Survey of students PAPI (in-school); absent students, school and class changers, school-leavers PAPI, CATI or CAWI (in-home) Wave 3: PAPI, CATI, or CAWI (in-home)Wave 4: PAPI, CATI, or CAWI (in-home) Wave 5: PAPI, CATI, or CAWI (in-home) Wave 6: CAPI (face-to-face); otherwise PAPI, CATI, or CAWI (in-home) Wave 7: PAPI, CATI, or CAWI (in-home) Wave 8: PAPI, CATI, or CAWI (in-home) Wave 9: PAPI, CATI, or CAWI (in-home) Detailed information can be found on the CILS4EU website. Administrative data: Identification of the respondents in the Integrated Employment Biographies of the IAB
Institutions involved	Survey data Contracting body: Mannheim Centre for European Social Research (MZES), University of Mannheim (funding: NORFACE ERA NET Plus Migration in Europe-programme; DFG long-term funding) Implementation of the survey

Category	Description
	IEA Data Processing Center DPC; Mannheim Centre for European Social Research (MZES), University of Mannheim; infas Institute for Applied Social Sciences Administrative data Social security agencies, Federal Employment Agency (Bundesagentur für Arbeit), municipal institutions
Frequency of data collection	Survey data: Waves 1-6 annually; later waves: biennial Administrative data Continuous
File format and size	Survey data: CILS4EU_DE (Stata): 30 data files between 44 KB and 47 MB Administrative data: Stata; Individual File: 10 MB; Basis Establishment File: 2 MB
File architecture	Survey data: The data is stored in several files (wave- or episode-specific). Administrative data: The data are stored in two files. One contains individual-level information and the other establishment-related information. Further files with additional information on establishments can be provided following a justified application. Linkage file: File containing the person ID (youthid) as well as information on the matching success
Data access	On-site use or remote data execution
Degree of anonymisation	Survey data: Pseudonymised Administrative data: Weakly anonymous
Sensitive variables	Month of birth (gebmon), Nationality (nation), measure type – group (mass), occupational sub-group (beruf2010_4), place of residence: employment agency (wo_aa), place of residence: district (Kreis) (wo_kreis) place of work: district (Kreis) (ao_kreis), economic activity - sub-class of economic activity (five-digit code) (w08_5).
Citation of data and data documentation	**Data**: Antoni, Manfred; Arnold, Lena; Bauer, Marlene; Dollmann, Jörg; Graf, Tobias; Grießemer, Stephan; Kalter, Frank; Kogan, Irena; Köhler, Markus; Lehnert, Claudia; Moczall, Andreas; Oertel, Martina; Schmucker, Alexandra; Schneider, Andreas; Soiné, Hannah; Weißmann, Markus (2025): "Children of Immigrants Longitudinal Survey in Four European Countries - Germany linked to administrative data of the IAB (CILS4EU-DE-ADIAB) – Version 7523 v1". Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). DOI: 10.5164/IAB.CILS4EU-DE-ADIAB7523.de.en.v1

Category	Description
	The data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and remote execution. **Data documentation**: Antoni, Manfred; Dollmann, Jörg; Schmucker, Alexandra; Weißmann, Markus (2025): Children of Immigrants Longitudinal Survey in Four European Countries - Germany linked to administrative data of the IAB (CILS4EU-DE-ADIAB) 1975-2023. FDZ-Datenreport, 05/2025 (en), Nürnberg. DOI: 10.5164/IAB.FDZD.2505.en.v1 **Further information on CILS4EU and CILS4EU-DE data can be found:** Kalter, Frank, Kogan, Irena, Jörg Dollmann (2019): Studying Integration from Adolescence to Early Adulthood: Design, Content, and Research Potential of the CILS4EU-DE Data. European Sociological Review, 35(2), 280–297. https://doi.org/10.1093/esr/jcy051 Especially, the sampling and the first survey of both subsamples are described in the technical reports of wave 1 and 6: CILS4EU. 2016. Children of Immigrants Longitudinal Survey in Four European Countries. Technical Report. Wave 1 – 2010/2011, v1.2.0. Mannheim: Mannheim University. Schiel, Stefan, Katharina Sandbrink, Folkert Aust, Nina Chudziak, and Martin Kleudgen. 2016. Methodenbericht. CILS4EU-Projekt: CAPI-Befragung von jungen Erwachsenen in Deutschland. Bonn: infas Institut für angewandte Sozialwissenschaft GmbH.
Dataset version	Children of Immigrants Longitudinal Survey in Four European Countries - Germany linked to administrative data of the IAB (CILS4EU-DE-ADIAB) – Version 7523 v1; DOI: 10.5164/IAB.CILS4EU-DE-ADIAB7523.de.en.v1

1.5 Volume structure

Table 2: Volume structure

No. of cases	before splitting	after splitting
BeH	50,578	66,560
LeH	2,273	3,900
LHG	5,840	13,312
ASU	10,761	19,670
XASU	1,268	2,404
MTH	1,978	3,411
XMTH	134	293
Total number of observations	72,832	109,550
Individuals	3,594	3,594

2 Data sources

2.1 Children of Immigrants Longitudinal Survey in Four European Countries – Germany

The CILS4EU(-DE) project provides a comprehensive dataset that enables the analysis of the complex causal processes underlying the integration of immigrant children and children with a migration background during their transition from adolescence to adulthood. It focuses on the structural, social, cultural, and emotional dimensions of integration. To this end, the study regularly collects information about their educational attainment, friendship networks, attitudes, identity, and satisfaction with various life domains.

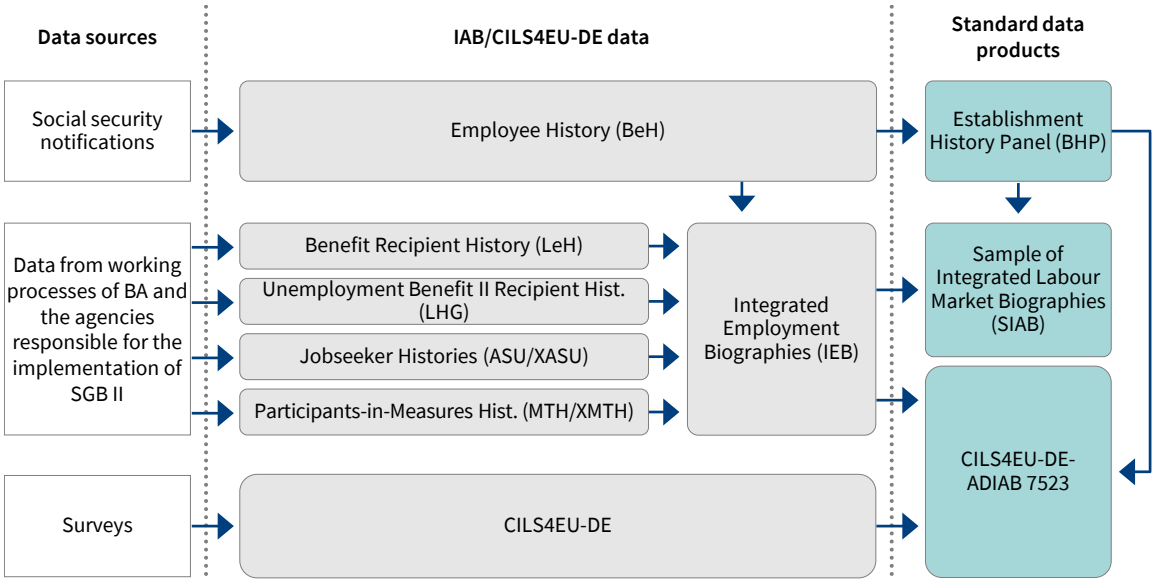
Beyond these core domains, additional topics such as ego-centric friendship networks (Waves 1, 3, and 7), information on siblings (Waves 4 and 8), political and social participation (Wave 5), and romantic relationships (Waves 4, 6, and 9) were addressed in wave-specific in-depth modules. Furthermore, the data was complemented by cognitive and language achievement tests (Waves 1 and 6), sociometric surveys of class networks (Waves 1 and 2), and analyses of speech recordings of respondents in terms of foreign and regional accents (Wave 6).

The dataset therefore allows for the analysis of various integration processes, such as educational choices and outcomes, the development of friendships and romantic relationships with persons with and without migration background, or changes in national and ethnic identity or religiosity over the course of coming of age. Moreover, it is possible to relate these processes to one another in order to account for the complex interaction of the different dimensions.

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 in accordance with the Sample of Integrated Labour Market Biographies (SIAB, see Schmucker et al., 2025). They unite data from different data sources, each of which may contain information from different administrative procedures. In addition, some supplementary variables from these data sources, which are not part of the IEB, are incorporated into the administrative individual data. Figure 2 illustrates the data flows that lead to the CILS4EU-DE-ADIAB and their relationship to other FDZ data products.

Figure 2: Data sources of CILS4EU-DE-ADIAB



The oldest sources of the IEB date back to 1975. The following descriptions of the administrative data sources and the variables generated from them for CILS4EU-DE-ADIAB therefore also contain information on periods that are not covered in the CILS4EU-DE-ADIAB data due to the relatively young age of the participants in the survey. The employment histories in the CILS4EU-DE-ADIAB data begin in 2007 at the earliest. Accordingly, establishment data is also only available from 2007 onwards. Due to this special aspect, variables on the economic activity classifications WZ73, WZ93 and WZ03 are not included, even though they are available in other FDZ data products.

2.3 Employee History (BeH)

The source of data regarding employment is the Employee History (Beschäftigtenhistorik - BeH) of the IAB. The data basis is the integrated notification procedure for health, pension and unemployment insurance, which came into effect as of 1 January 1973 (and was extended to cover East Germany as of 1 January 1991) and is known by the abbreviation DEÜV (previously DEVO / DÜVO) (for further details see: Bender et al. 1996, p. 4 et seq.; Wermter /Cramer 1988). Under this procedure, employers are required to submit notifications to the responsible social security agencies concerning all of their employees covered by social security at least once a year. The BeH covers all white- and blue-collar workers as well as apprentices as long as they are not exempt from social security contributions. This means that civil servants, self-employed persons and regular students[2] (see Cramer 1985) are not recorded in the BeH in principle. Since the notification procedure was changed on 1 January 1999, employees in marginal part-time employment and unpaid family workers have also been recorded (not contained in the data until 1 April 1999). The data are recorded by the health insurance companies, collected and edited by the Federal Employment Agency (BA) and subsequently integrated into the History File by the IAB.

The employment episodes in the administrative individual data are supplemented by establishment data (Basis Establishment File and BHP Extension Files). They are taken from the

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

Establishment History Panel (Betriebs-Historik-Panel - BHP), which is also based on the BeH. The Basic Establishment File and BHP Extension Files, which are provided with the CILS4EU-DE-ADIAB data, comprise only information on establishments from which at least one employee is included in the CILS4EU-DE-ADIAB sample. Furthermore, the establishment data are limited to the years in which the employees appear in the administrative personal data.

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

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

Box 1: Example code for merging individual and establishment data in Stata 17

```
use $orig/CILS4EU-DE-ADIAB_7523_v1.dta
gen int jahr = year(begepi)
sort betnr_cils4eu jahr
merge m:1 betnr_cils4eu jahr using $orig/CILS4EU-DE-ADIAB_7523_v1_bhp_basis_v1.dta
```

2.4 Benefit Recipient History (LeH)

The Benefit Recipient History (Leistungsempfänger-Historik - LeH) of the IAB covers periods during which individuals receive earnings replacement benefits from the Federal Employment Agency (sphere of Social Code Book III). The benefits comprise unemployment benefit, unemployment assistance, maintenance allowance, and contributions paid by the BA to private health or care insurance while benefits are being drawn. Furthermore, information on suspension and interruption periods is also included:

- A suspension period is a period during which a person who is entitled to benefits does not receive them for reasons he or she is responsible for. A typical example is a short suspension period at the beginning of a benefit claim, if the person registered as a jobseeker too late.
- An interruption period is a period without receipt of benefits for reasons the person entitled to benefits is not responsible for. This might be, for example, a compulsory absence from the place of residence. A period of interruption does not reduce the duration of entitlement, i.e. the person does not receive any benefit during the period of interruption, but the regular end of entitlement to benefits is extended accordingly.

Benefits in the context of Social Code Book II (e.g., unemployment benefit II) are not included in the data. Since entitlement to benefits depends on meeting certain legal requirements, periods of

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

unemployment in which the requirements are not met (e.g., no eligibility for unemployment assistance, or non-completion of the qualifying period for unemployment benefit) are not reported in the Benefit Recipient History. The earliest available data in the LeH are from 1 January 1975.

2.5 Unemployment Benefit II Recipient History (LHG)

The Unemployment Benefit II Recipient History (Leistungshistorik Grundsicherung - LHG) contains information about individuals who are eligible for benefits and capable of work, about the members of their benefit unit (Bedarfsgemeinschaft) in accordance with § 7 SGB II and about certain individuals associated with the benefit unit. In CILS4EU-DE-ADIAB, however, it is not possible to link individuals with benefit receipt in accordance with Social Code Book II (SGB II) within benefit units. The receipt of benefits in accordance with SGB II covers both basic social security benefits (e.g., Unemployment Benefit II) and supplements to unemployment benefit or additional benefits. The LHG does not contain any information about the benefit rates, however. As the amount of benefit received is not determined at the level of the individual but at the level of the benefit unit in the case of Unemployment Benefit II, it is difficult to assign an individual benefit rate. Therefore, CILS4EU-DE-ADIAB also does not contain information about SGB-II-benefit rates.

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

- Cooperation of employment agencies and municipalities (Arbeitsgemeinschaften – ARGE) until the end of 2010 / joint facilities (gemeinsame Einrichtungen) since 2011), in which the BA and the municipality deal with tasks jointly,
- separated responsibilities (getrennte Trägerschaft - gt) / municipalities exercising their duties separately (until 2011) – here the tasks are divided between the BA and the municipality,[4]
- authorised municipalities, which are also called opting local authorities or opting municipalities according to the initial experimental clause of Section 6a - here the local authority is responsible for all tasks in the sphere of SGB II.

The data of the “Unemployment Benefit II Recipient History drawn from A2LL” (LHG) come from different reporting procedures. As a rule, the IT procedure A2LL was used in all ARGE cooperation projects until 2010, and in joint facilities from 2011 onwards.[5] Authorised municipalities use various IT procedures of their own and transmit their data to the BA by means of the XSozial-BA-SGB II standard once a month. Both of the procedures are used by municipalities with separated responsibilities. The different data standards affect the scope and quality of the data supplied.

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

[4] 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.

[5] 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.

2.6 Jobseeker Histories (ASU/XASU)

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

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

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

3 Data preparation and sampling procedure

3.1 Corrections and validation procedures

3.1.1 Survey data CILS4EU-DE

Detailed information on the sampling design, data collection procedures, survey instruments, and data structure of CILS4EU and CILS4EU-DE is available in the documentation provided on the CILS4EU website (see https://www.cils4.eu under “Documentation”). Information on the sampling design of the two subsamples can be found in the technical reports for Wave 1 (CILS4EU, 2016) and Wave 6 (Schiel et al., 2016). Furthermore, codebooks for each wave provide information on the respective datasets as well as on the variables and their values.

For CILS4EU-DE-ADIAB, the original identifiers of the CILS4EU-DE survey data (youthid, classid and schoolid) have been replaced with new identifiers for data protection reasons. However, the variable names remain unchanged.

3.1.2 Administrative data from the Integrated Employment Biographies (IEB)

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

- Observations in which the age is under 13 or over 75 are deleted.

- Observations whose end date precedes the start date are deleted.
- Inconsistent information on gender or date of birth within an account is corrected.
- Records with no information on the date of birth or on gender after the correction procedure are deleted.

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

3.1.3 Employee History (BeH)

- To capture a person group that is as constant as possible over time, some person groups for which data are not available throughout the entire observation period are excluded. From the reporting year 2011 onwards, the BeH data originate from newly designed source data. As a result, a number of person groups have been introduced or reactivated as they are classified by the BA statistics as being subject to social security contributions. The person groups 101 - 107, 111 - 114, 118, 119, 120, 140, 141, 142, 143, 149, 201 and 203 - 205 are therefore contained from that time onwards as well as the two groups 109 and 209, which indicate people in marginal part-time employment. Groups that are not included are, for example, people in short-term employment, i.e. person groups 110, 202 and 210.
- Person groups 123, 124 and 127 have been newly introduced in 2011.
- For data protection reasons, the person groups 107, 111, 113, 114, 127 and 204 are combined to form the person group “other workers” (599).
- From the reporting year 2012 onwards apprentices were included as the new person groups 121 and 122.
- Observations with no details on earnings, and the value 101 for the person group variable, and the value 50 for the reason for notification (annual notification) are not incorporated into the IEB.
- Gender and date of birth are taken from the Data Warehouse (DWH) of the BA. This information is harmonised across data sources.
- The territorial allocations for place of work and place of residence are updated to the status as of 31 December 2023.

3.1.4 Benefit Recipient History (LeH)

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

3.1.5 Unemployment Benefit II Recipient History (LHG)

- Observations without a BA client number are deleted.
- Cancelled data records are not used.
- It only contains observations of people who are capable of work and people under the age of 65.
- In each case, non-overlapping periods of benefit entitlement of a person in a certain benefit unit (BG) are created. New observations are begun for the following administrative reasons:
  - on certain birthdays of members of the BG that are stipulated by law and relevant for structural changes in the benefit unit (14, 15, 18 and 25) and the individual retirement age of members of the BG (see Section 3.1.8),
  - when the structure of the benefit unit changes (e.g. due to entries/exits),
  - when there are changes in a variable of the BG client and
  - at the beginning and the end of a case of benefit sanctions for observations from 1 April 2006 onwards. It must be taken into account, however, that it is not possible to identify the duration or type of sanction or the time when it was imposed or when it began on the basis of the data. The reason for this is that there is no corresponding variable or value that indicates the start, type or duration of the sanction.
- For the reason mentioned above, all individual-related variables that are available for the LHG source are valid for the entire duration of the observation.
- Double notifications due to the territorial reforms in 2009/2011 and the reorganisation of the institutions in 2012 (see Section 5.2.2) were corrected as far as possible.
- The territorial allocations are corrected in the same way as in the BeH.

3.1.6 Jobseeker Histories (ASU/XASU)

- Observations with an end date before 1 January 1995 are not included.
- There is no consolidation of the ASU observations for individual persons. Therefore, overlaps between ASU observations might occur.
- Individual-related variables that are only available for the (X)ASU sources always refer to the beginning of the spell.
- A new ASU spell is generated as soon as a change of status occurs (e.g., from seeking work to unemployed). This also applies if the type of institution (employment agency, cooperation of employment agency and municipality, joint facility, authorised municipalities, separated responsibilities) changes.
- The XASU contains non-overlapping time periods for individuals. If one of the following variables changes, in each case a new data spell is generated for the XASU:
  - change of job search status
  - change of availability
  - change of SGB II institution (due to notification procedure)
  - change of place of residence
- The territorial allocations are corrected in the same way as in the BeH.

3.1.7 Participants-In-Measures History File (MTH)

- Observations generated more than a year after the end of the measure are deleted if another observation exists that was generated within the year after completion of the measure.
- Only the most recent record of an individual case of participation in a measure is used.
- Only cases of participation in measures that are classified as "actually took place" are included in the IEB. Cases of participation that did not take place or have not yet taken place are deleted. Cases of participation are also classed as not having taken place when a deletion date is set during the participation in a measure.
- Certain types of measure are not included. These include services to support careers advice and job placement, mobility assistance and pure rehabilitation measures.
- The territorial allocations are corrected in the same way as in the BeH.

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

- For XMTH, the particular challenge is to identify and handle multiple notifications of participation that are caused by technical or organisational issues. The causes of these duplicate notifications are, on the one hand, a missing identification number for participation in measures until April 2009 or improper handling of the subsequently introduced promotion ID and, on the other hand, a change of provider numbers in connection with mergers or other reorganisations of institutions responsible for implementing SGB II due to the reorganisation of the employment agencies in 2012/2013. In addition, provider-specific problems arose with version or product changes of the municipal software or with version changes of the XSozial standard. At times, this results in a considerable number of reports of participations in measures of the same type, which partly or completely overlap in time per participant. Real duplicates are identified and sorted out because they are highly likely to be technical duplicates. Hidden duplicates can also be identified and sorted out. In principle, the rule always applies that the most up-to-date information is retained.
- In addition, notifications with a starting date before 2005 are excluded.
- Overlapping and immediately adjacent notifications of the same type of measure (from XSozial) are combined to one observation. Measure-specific characteristics of the combined periods are set to system missings.
- Subsequently, the remaining characteristics are compiled and calculated. For this, consolidated individual data from XSozial-histories and LHG are used.
- Finally, the following measures are excluded:
  - one-off benefits (like UBV/Mobi-/Vermittlungsbudget /LES) and
  - specific rehabilitation measures
- The territorial allocations are corrected in the same way as in the BeH.

3.1.9 SGB II anonymisation

In order to reduce the risk of de-anonymisation, only the year of birth is available in CILS4EU-DE-ADIAB by default. The month of birth can only be requested as a sensitive variable if there is a justified need for it. However, in the LHG and (X)ASU there is the risk that the exact date of birth

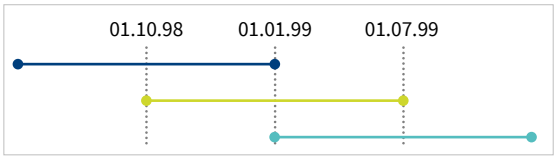
may still be obvious due to the chronological structure of the observations. Observations might end systematically on certain birthdays and/or the day before, or start again on the birthday.

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

3.2 Episode splitting

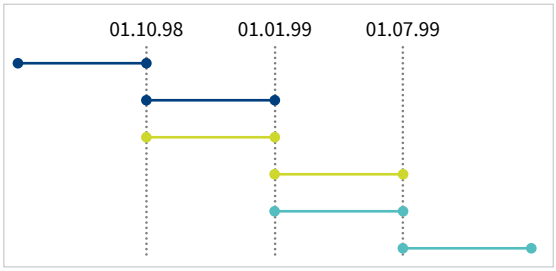
The administrative individual data are available with “split” episodes. If observations overlap within an account, these observations are replaced by artificial observations with new dates so that completely parallel periods and non-overlapping periods are created. This increases the number of observations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

Figure 3: Episode splitting



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

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



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

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

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

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

Box 2: Example code to create additional observation counters in Stata 17

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

3.3 Missing values

In the SIAB, missing values are coded as follows:

Term	Value	Description
No (valid) details available	.z	Values of a variable that are not systematically missing, i.e., the variable is available in principle for the data source, but no details are available for the value considered or cannot be interpreted reasonably.
Systematically not available	.n	A variable is not available in principle for a data source 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 CILS4EU-DE. 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 (example for the question in wave 2022):

"First of all, thank you very much for your longtime participation in the Youth in Europe Study (YES) and your willingness to take part in the survey again this year.

When evaluating the survey, we would like to include excerpts from data that the Institute for Employment Research at the Federal Employment Agency in Nuremberg has on you. This includes, for example, information on employment relations, periods of unemployment, and company characteristics. Your data from previous YES surveys are less valuable if we cannot supplement them with further data.

For the purpose of adding these data to the interview data, we would like to ask you cordially for your consent. It is absolutely guaranteed that all data protection regulations will be strictly adhered to. Your consent is of course voluntary. You can withdraw it at any time. If you refuse, you will not suffer any disadvantages.

Do you agree to this?"

Respondents who did not agree were excluded from the sample for the linkage. The consent question was first asked in Wave 9 (2022) and was repeated in Wave 10 (2024) for respondents for whom no consent was obtained in Wave 9.

Table 3 provides an overview of how many respondents were asked by CILS4EU-DE to consent to data linking up to and including 2024 and how many gave their consent. If individuals were asked for their consent to linking multiple times over the course of time and gave different answers (consent/refusal), the later or most recent answer was used as the basis for the present data product. This means that only individuals who agreed to the linking of their data and did not withdraw their consent at a later date were taken into account. The consent rate for data linking was 90.5%.

Table 3: Requested and given consent for data linkage in CILS4EU-DE

Groups in CILS4EU-DE	Frequencies or Share
Asked for consent	4,043
Consent given (until August 2024)	3,725
Share	92.1
Excluded from linkage due to implausible information	65
Consent withdrawn	1
Consent revised	3,659
Share revised	90.5%

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 CILS4EU-DE who had given their consent:

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

To identify the respondents in the administrative data of the IAB, an excerpt from the central address files of the Data Warehouse of the Federal Employment Agency was used. These address files are restricted to the birth years surveyed in CILS4EU-DE and the postcode areas derived from the sampling points.

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

4.3 Correction and standardisation of personal information

The above-mentioned personal information from CILS4EU-DE and the administrative data was subjected to correction routines before the matching in order to achieve the best possible correspondence between the two data sources. For this so-called pre-processing, established procedures were used which were refined by the German Record Linkage Center (GRLC) at the IAB-RDC for the peculiarities of the German context (Antoni/Schnell, 2019 or https://www.record-linkage.de). See Christen (2012) for a comprehensive overview of all the methods described in Sections 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). 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. The next step was to find out how many of the remaining (unlinked) people could be linked if a discrepancy in individual address fields or the date of birth was tolerated (while the first and last name fields still had to match exactly) applying an iterative process. 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 successively allowed in, respectively, of one of the particularly error-prone items: house number, postal code, town, day and month of birth, street or year of birth.

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 nonetheless 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 ToolBox (MTB, see Schnell et al.,

2004). For the first and last name as well as for the town and the streets a string similarity was calculated by means of the Jaro Winkler metric. 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, an iterative blocking strategy was applied. First, the observations were grouped into blocks at the level of the five-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 five-digit postal code. In addition, this reduces the risk of linking different people with very similar identifiers. However, it can happen to a limited extent that no correspondence can be found for individuals in the administrative data - for example due to missing or incorrect information about the postal code or because a move to another postal code area is only recorded in one of the two data sources. Thus, an additional linkage was performed using gender, year and month of birth as well as first name, three-digit postal code and year of birth in each case as blocks (see Table 4).

Table 4 Iterative blocking strategy

Step	Blocking variable(s)
1	Postal code (5-digit)
2	Gender, year and month of birth
3	First name
4	Postal code (3-digit)
5	Year of birth

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

Of all the compared pairs of records, the corresponding match in the IEB data with the highest similarity value was selected for each respondent and checked manually. Only plausible matches were kept and implausible matches were removed. Unambiguous cases were marked as potential matches for closer examination later. For the remaining individuals who could not be assigned a suitable match, this procedure was repeated in the subsequent blocking step.

Finally, a manual linkage was performed for all persons with a potential match, for whom no suitable match could be found by any of the above mentioned steps. For this purpose, all existing

entries of the persons collected in the previous steps were inspected in both data sources and plausible matches were adopted.

Due to the manual checks, the probability of identifying different persons as a match during the error-tolerant linkage was extremely low.

4.6 Result of the data linkage

It was possible to use data of 3,659 of these people for the data linkage. For 3,650 persons, the survey data could be linked to an account in the administrative data. This corresponds to a very high linkage rate of 99.8%. The linkage rate refers to the people available for data linkage. Only 9 individuals could not be found in the administrative data, despite having given their consent and being available for the linkage. Table 5 summarizes the success rates of the individual linkage steps.

Table 5 Results of the record linkage

Group of persons	N	Percantage (in %)
Respondents providing consent for linkage	3,659	100,0
Matched respondents	3,650	99.8
Matches with IEB	3,594	98.2
Including:		
Exaktes matches	3,611	98.7
Probabilistic matches	35	1.0
Manuel matches	4	0.2

5 Data quality and problems

5.1 CILS4EU-DE

A description of the data as well as information on the sampling procedure and data collection can be found in the documentation of the CILS4EU and CILS4EU-DE data, provided on the CILS4EU website (see https://www.cils4.eu).

For any questions or problems, please contact the project team at the Mannheim Centre for European Social Research (MZES)/University of Mannheim at cils4eu@mzes.uni-mannheim.de.

5.2 Entire IEB

5.2.1 Gaps in employment histories

The IEB contains employment histories. However, not every type of employment is included in the administrative data. Some individuals with certain life courses are not represented in the IEB at all.

For evaluation purposes, it is often relevant to know gaps in the included biographies (e.g. for creating control groups, analysing life courses, etc.). The gaps listed below are defined as periods of time after the end of school education for which no data are included in the IEB. These gaps can be divided into

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

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

Table 6: Biographical gaps and possible ways of identifying them

Biographical gap	**Information on gap, potentially identifiable using the details in the "grund" variable in the preceding observation of the source**
Civil servants, professional soldiers, judges, employees of bodies or foundations under public law	XASU
Self-employed persons without support	LeH, ASU
Students, persons in school-based further education	LeH, LHG, ASU, XASU
Persons who are ill / not able to work for more than 6 weeks (illness during unemployment, however, is represented in the ASU source under certain circumstances, see Section 3.1.6)	BeH, LeH, ASU
Persons receiving old-age pension without employment if not a member of a benefit unit	LeH, LHG, ASU
Individuals on maternity leave / parental leave	XASU
Recipients of early retirement benefits	LeH, ASU
Trade professionals working from home	
Employees working short-time	ASU
Persons in youth welfare facilities, in vocational training centres, approved workshops or similar facilities for people with disabilities	ASU
Participants in programmes to support participation in working life (people in rehabilitation)	ASU
(Sideline) farmers	
Caregivers according to Section 19 SBG XI	
Conscripts	BeH, LeH, LHG, ASU, XASU
Persons in reserve duty training	BeH, LeH, LHG, ASU, XASU
Persons fulfilling community service	BeH, LeH, LHG, ASU, XASU
Persons fulfilling a voluntary social or ecological year instead of community service	
Other people not registered with the statutory pension insurance or the Federal Employment Agency (e.g.	BeH, LeH, ASU

Biographical gap	Information on gap, potentially identifiable using the details in the "grund" variable in the preceding observation of the source
sabbatical, funding from personal assets or pensions, emigration, employment abroad, voluntary work etc.)	
Strikers in cases where the strike lasts more than a month	LeH
Social assistance recipients (prior to the introduction of SGB II in 2005), recipients of welfare payments (according to SGB II)	
SGB-II recipients whose providers have experienced delivery failures	
Recipients of compensation according to FELEG (Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit, Act on Support in Case of Termination of Farming Activities)	

5.2.2 Introduction of SGB II and subsequent institutional changes

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

- Usually, the Arbeitsgemeinschaften (ARGE) between BA and the district took over the tasks according to the SGB II. These were replaced in 2011 by the joint facilities (gE) - also known as "Job Centers".
- In addition, it was possible until the end of 2011 for the BA and the district to perform the tasks assigned to them in their own (separate) responsibility (gT and gAw).
- Since 2005, it has also been possible for authorised municipalities (zkT; also: opting municipalities) to take over the tasks. Initially, 69 administrative districts or independent towns took sole responsibility for the basic provision for job seekers until 31.12.2010. With a constitutional amendment, the option was extended to 110 providers from 2012 onwards.

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

As part of the reorganisation of SGB II institutions on 1 January of 2011 and of 2014, various ARGEn/gE and separate providers were transferred to municipal providers or converted from zkT to gE. This change of responsibility and the associated change of the software used locally also

lead to some breaks in the data of the SGB-II sources. Detailed information on the resulting quality problems in the data can be found in the individual sections on the data sources.

5.3 Employee History (BeH)

- In a few cases (less than 1%), the employee history contains duplicates in the variables individual ID, establishment ID, employment status, original start date and original end date. The notification procedure does not provide for such duplicates. In most cases, the individual records in a duplicate group differ in the daily wage, in rarer cases also in other variables. The reason for such duplicates is probably incorrectly processed cancellations.
- Information on vocational training, the occupation/activity performed and the occupational status is transmitted by means of notifications made by the employer in accordance with the Data Collection and Transmission Regulation (DEÜV) (see Section 2.3) using a so-called occupation code. The new occupation code 2010 was adopted for notifications with an end date later than 30 November 2011 (for further details, see Bertat et al., 2013). The decision to switch to the new occupation code was made by the central organisations of the social security agencies as a number of facts could no longer be recorded in a way that was up-to-date and realistic using the occupation code 2003. As the notifications made by employers in accordance with DEÜV only enter the Employee History (BeH), the change of the occupation code only affects observations from this source. The measurement of the following characteristics previously reported using the occupation code 2003 is affected by this change: working hours, occupation, occupational status and school and vocational qualification level. In addition, since the switch to the new occupation code, details are also available about whether an employment relationship is fixed-term and whether a person is employed by a temporary work agency to be hired out to other firms. The most important consequence is the switch to a new occupational classification. Instead of the previous Classification of Occupations 1988 (Klassifikation der Berufe 1988 (KldB 1988)), the more highly differentiated KldB 2010 is reported with the new occupation code.[6]
- The variable "Occupational status and working hours" (stib) is only filled for reports that were submitted before the introduction of the new occupation code. The FDZ does not extrapolate or impute for later reports. The categories of the variable stib that can be consistently observed over the whole observation period (occupational status as a trainee, distinction between part-time and full-time) are to be reproduced in the variables erwstat and teilzeit.
- The introduction of the new occupation code in 2011 led to a number of problems. For example, during the transition period granted to employers in the social security notification procedure,[7] there was a temporary increase in the number of missing details. Analyses of the BA statistics (Bertat et al., 2013, p. 10) show that in 20 to 30 percent of cases no information was available in the new or converted variables “occupation/activity performed”, “working time” and “vocational education and training” after the switch. This situation began to improve significantly in the first half of 2013. In order to improve the quality of the “working time” variable in the transition period, Ludsteck/Thomsen (2016) developed an imputation

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

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

procedure to replace the missing values by imputed values. The imputed data are included in the SIAB 7523. No imputation is performed regarding the gaps in the other variables.

- Due to the introduction of the employment notification procedure in the federal states of eastern Germany, the notifications for eastern Germany can only be assumed to be sufficiently complete from 1993 onwards. For the same reason, a large number of spells for 1991 have missing values for several variables (such as vocational training, employment status, and daily wage).
- The increase in the number of BeH observations from 1999 onwards is due to the introduction of the obligation to submit employment notifications for people in marginal part-time employment from 1 April 1999 onwards.
- Especially in 1999, observations of part-time employment increase significantly. This is caused by the actually observed increase in part-time work as well as by the fact that since 1999 employment notifications have been completed more correctly.
- Within the employment notification procedure, a certain time lag is unavoidable. Although changes in employment relationships have to be reported immediately, and existing employment relationships have to be confirmed annually by April (until the annual notification 2012) or mid-February (from the annual notification 2013 onwards) of the following year, some notifications actually arrive years later. The History File of the IAB is not updated continuously, however, but at certain intervals. This is done using files of employment notifications for one particular year which were submitted 36, 18, 12 or 6 months after the end of the reporting year (e.g. the 18-month file for 2013 can be created in July 2015 at the earliest). Notifications submitted more than three years late are not taken into account at the IAB, which means that a 36-month file shows a 100 % degree of completeness by definition.
- In the version of the IEB on which the SIAB 7523 is based, the year 2020 is the last year with a degree of completeness of BeH observations of 100%. For the years 2021 and 2022, the 18-month files were used, and the observations for 2023 originate from a 6-month file. It can therefore be assumed that employment notifications for 2021 and 2022 are slightly underreported in the SIAB, and that those for 2023 are underreported to a slightly higher degree. However, this should not reduce the ability to analyse the data at individual level. The missing notifications occur more frequently in a few establishments, however. This means that in individual cases the establishment data, e.g. establishment size, are grossly incorrect and will change considerably in subsequent versions.
- In 1984, a change was made in the employment notification procedure. From that time onwards, one-off payments of gross earned income were reported as part of the annual earnings subject to social security contributions, which leads to an increase in the average daily wage. In particular, the proportion of wages and salaries above the upper earnings limit increased considerably from that year onwards (see Bender et al. 1996).
- For the years 1992 until 2000, noticeable decreases and increases in the number of notifications were observed. Decreases can be observed especially for the following 10 districts: Braunschweig (03101), Wolfsburg (03103), Emden (03402), Kassel (06633), Essen (05113), Neuss (05162), Erftkreis (05362), Hersfeld-Rotenburg (06632), Miltenberg (09676) and

Kempten (Allgäu) (09763). This is due to notification problems of one or more establishments in these regions.

- Considerable decreases were also observed for the districts Salzgitter (03102) and Hoyerswerda (14264).
- Concerning the notifications for full-time employment, especially the districts Main-Taunus (06436) and Alzey-Worms (07331) are noteworthy. They feature above-average rises. Also in this case, the reasons are notification problems at one or more establishments in these regions.
- In the years 1996 to 1998, the values 841-844 (doctors and pharmacies) within the 'occupation – activity performed' variable are very rare compared to the neighbouring years. The reasons for this are not known.
- In the years 1975 and 1977, there were so far considerably too many employees with a place of work municipality of Lahn (district of Emsland, Lower Saxony). Instead of the expected double-digit figure, there were up to 90,000 employee registrations per year with this place of work in the population of the BeH. The reason for this is a historical misclassification of employment reports from the city of Lahn. This was a merger of the Hessian municipalities Wetzlar, Gießen, Heuchelheim, Wettenberg and Lahnau (districts of Gießen and Lahn-Dill-Kreis) which was dissolved after a short time. On the basis of these findings and assuming that there were no real company relocations between these regions, the following adjustment rule was implemented at the level of the company number: As soon as the company location changes from the municipality of Lahn to one of the listed Hessian municipalities in the years 1975 to 1978, the former specification of the municipality of Lahn is overwritten with the later correct specification. This rule significantly reduces the overhang and the municipality of Lahn in Lower Saxony then only has less than 2,000 employees in the population of the BeH in the critical years.

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 benefit suspension and interruption in 2004. Until 1 July 2004, such periods 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 these periods

5.5 Unemployment Benefit II Recipient History (LHG)

- With regard to the completeness of case numbers or benefit histories from the LHG data sources, there are substantial gaps in the years 2005 and 2006. We therefore strongly advise against analysing the data for this time period based merely on the LHG sources.

- Longitudinal analyses of individuals are affected by inaccuracies as it is not possible to distinguish between changes in the benefit entitlement status and relocations into and out of districts whose institutions had problems delivering data.
- Also from 2007 onwards, cases of underrecording occur at times. These generally last one month and occur mainly in the authorised municipalities.
- Underrecording and overrecording occur in connection with changes in the type of institution responsible for implementing SGB II:
  - In the context of the reform of the territories covered by the institutions, which came into force on 1 January 2011, cases of underreporting occurred in the districts covered by the employment agencies of Dessau-Roßlau, Halberstadt, Halle and Sangerhausen.
  - Double notifications due to the territorial reforms in 2009/2011 and the changes in the form of the institutions as of 1 January 2012 and of 1 January 2013 are already corrected as far as possible in the IEB. Nonetheless double notifications may still occur.
- In the following job centres there are inaccuracies with regard to the allocation of benefit cases:
  - between Emden and Norden between September and December 2009
  - between Döbeln and Mittelsachsen from October to December 2012
  - between Tirschenreuth and Wunsiedel from November 2012 to March 2013
- Some individuals for whom a (X)LHG spell exists are excluded entirely or partly from benefit receipt according to SGB II, for instance because they take part in a subsidised training programme, receive an old-age pension, live in an in-patient facility or a residential institution or receive insurance payments aimed at avoiding need. This affects on average 3 to 5 percent of all cases. In XSozial this person group is sometimes underrecorded by some institutions. Exclusion from benefits cannot be identified in CILS4EU-DE-ADIAB.
- Due to the reporting logic, information from the XSozial transmission standard can only be updated monthly.
- In the official performance statistics of the BA, reporting gaps are supplemented by a statistical estimation procedure at an aggregated regional level. No supplementary data sets are provided in the LHG. A comparison of the IEB with the performance statistics is therefore only possible to a limited extent.

5.6 Jobseeker Histories (ASU/XASU)

5.6.1 ASU

- The registered periods of job search activity in the ASU source are regarded as complete from the year 1997 onwards. Therefore, the analysis potential of the ASU spells before 1997 is limited.
- With the introduction of SGB II on 1 January 2005, jobseekers are no longer fully covered by BA procedures. From this date, the ASU only covers persons who are supported by the BA in the sphere of the SGB III (employment promotion) or by ARGEn, gE or gT in the sphere of the SGB II (basic security).

- For the placement staff it is not always possible to record the allocation to the legal sphere immediately, since it is frequently only clear which institution is primarily responsible after a certain time due to a possible entitlement to SGB II benefits. Therefore, we recommend comparing the value of the variable “Type of provider” in the ASU with the value in the LHG and/or XLHG for the same period of time. Due to the recording gaps in the LHG and XLHG between 2005 and 2006 this is not always possible.
- For some individuals for whom an authorised municipality has been responsible since 2005, parallel "artificial" ASU datasets were created by the Federal Employment Agency. These can be identified via the variable “Employment status prior to job search” (estatvor) (transfer to an authorised municipality).
- From mid-2005 until mid-2006, the coArb IT procedure, from which the jobseeker and applicant pool data originate, was superseded by the VerBIS procedure at the Federal Employment Agency. In July 2005, coArb was first replaced by VerBIS in the employment agency in Wiesbaden as a pilot project. From December 2005 onwards, it was then gradually replaced by VerBIS in several stages in all employment agencies. The information for many of the variables recorded was gathered with different levels of differentiation and different qualitative weighting in the two systems. It is therefore very difficult to integrate these variables into the IEB, which is only possible using a special procedure (mapping). Unfortunately, a full conversion of the affected variables from coArb to VerBIS cannot be achieved by means of mapping, so for some variables there is an unusually large number of the values ‘no details available’, ‘other’ or ‘missing’. Moreover, striking differences may occur in frequency counts, depending on whether the original source of the data was coArb or VerBIS. It can be assumed that integration agreements are under-recorded. Important limitations in the analysis potential are mentioned in the corresponding description of variables.
- The coArb procedure, which was used until June 2006, supported only the placement of unemployed persons and jobseekers. Some data were also collected about individuals who were only seeking advice, but these data are incomplete. The careers advice data were collected in a separate system. In VerBIS the attributes of the job-search status were extended to include ‘seeking advice’ and individuals ‘without status’. The latter group includes individuals eligible for Unemployment Benefit II who are only available for job placement to a limited degree. The recording of this group in VerBIS is only regarded as largely complete since January 2008.
- A change of the institution responsible for implementing SGB II or a change of place of residence does not lead to a new ASU observation, the value of the variable at the start of an episode is continued. The longer the observation becomes, the greater the risk is that the institution responsible or the place of residence is no longer correct.
- The job search status is hardly corrected afterwards, which is why the case numbers largely coincide with the BA statistics until 2005. Since spring 2011, the jobseeker data from BA procedures and XSozial have been consolidated in the integrated unemployment statistics. This may lead to larger deviations. At the current margin, however, the stock data are identical.

5.6.2 XASU

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

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

5.7.1 MTH

- The MTH is incomplete for measures with a start date before 1 January 2000.
- As of 1 January 2005, there is an inconsistency in the data as participants in measures were allocated to different institutions with the introduction of Social Code Book II (see Sections 2.7 and 5.2.2).
- The MTH only contains notifications that are recorded in BA procedures. The use of these procedures in cooperations of employment agencies and municipalities/separated responsibilities/municipalities exercising their duties separately increases continuously between 2005 and 2007. The notifications for these institutions are complete from March 2007 onwards. Measures that are reported by authorised municipalities via the XSozial standard are contained in the XMTH.

- Because of the reorganisation of the institutions responsible for implementing SGB-II in 2011 to 2014, a split of the documentation of participations in measures in the MTH and in the XMTH may occur when there is a change in the reporting procedure.[8] This might result in a split or a duplication of the spell of the measure (see Section 3.1.8).
- 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.

5.7.2 XMTH

- In the years 2005 to 2007 the notifications of participation in measures are incomplete. The degree of under-reporting is unclear due to a lack of or inaccurate comparative figures.
- Between 2005 and the beginning of 2017, about 13% of all institutions responsible for implementing SGB II report almost no municipal integration benefits (formerly accompanying benefits), which are included in the summarised category "other support". Many other job centers report only temporarily and/or only selected types. This leads to an under-recording of the "other support" category.
- Over-registration of participations: Total stocks are considered to be stable from the end of 2008. However, the majority of the providers still register notification profiles which are temporarily conspicuous and which raise doubts about the quality. Only the introduction of the XSozial promotion-ID in 2009 and the meanwhile several years of experience of all participants stabilise the reporting process and the subsequent data processing. Also the new zkT, introduced in2012, report inconspicuously for the most part, so that the scope and duration of presumed over-recording decrease noticeably.
- The reorganisation of SGB II providers in 2011-2014 (see Section 5.2.2) lead to a split in the documentation of participation in measures in MTH and XMTH if the reporting procedure was changed. This could result in a split of the measure spell, but also in duplications (see Section 3.1.8).
- The figures or persons received in the XMTH from monthly key date counts differ in several respects from the published BA statistics.

6 Description of variables

This chapter describes the variables contained in the administrative data in detail. Further detailed overviews of the individual variables (labels, counts, missing values) can be found in the

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

online appendix to the data report (see section 8.1). Further working tools can be found at https://fdz.iab.de/en/pd_hd/cils4eu-de-adiab-7523-v1-en/

6.1 Identifiers

6.1.1 Individual ID youthid)

Category	Description
Variable label	Individual ID
Variable name	youthid
Category	identifiers
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	The individual ID indicates which observations belong to the same person. It also indicates which persons correspond in survey and administrative data, i.e. youthid is included in both survey and administrative data. This identification variable is system-independent, which is why it does not allow any conclusions to be drawn about the characteristics of the person or original identifiers. For data protection reasons, the IDs available in CILS4EU-DE-ADIAB are not identical with the IDs in the reduced version of CILS4EU-DE (survey data) available for download, even though they have the same variable name.

6.1.2 Establishment ID (betnr_cils4eu)

Category	Description
Variable label	establishment ID
Variable name	betnr_cils4eu
Category	identifiers
Origin	BeH
Data type	numerical
Detailed description	The establishment ID indicates which observations belong to the same establishment. It is based on the establishment number allocated by the BA, which was replaced by an artificial number (further information on the allocation of establishment numbers by the BA can be found in Bender et al. (1996: p. 15 et seq. and pp. 27-30) as well as directly on the website of the establishment number service of the BA at https://www.arbeitsagentur.de/betriebsnummern-service/alles-wichtige). The establishment number and year specification can be used to merge individual and establishment information. For the establishment number, the following should be observed in general: If the company has only one office, or if the company has only one office in one municipality, this office is the establishment and is given an establishment number. If the company has **several** branch offices in **one** municipality, these establishment premises / workplaces must be merged into a single establishment under one establishment number, if they belong to the same economic class. If they do **not** belong to the same economic class, each branch office is regarded as a separate establishment and is given its own establishment number. If the company has **several** branch offices in **several** municipalities, each of these branch offices is an establishment and is given its own establishment number.

Category	Description
	In this context, the following definitions with regards to the allocation of establishment numbers as part of the notification procedure for social security must be observed: An **establishment** is a regionally and economically delimited unit in which employees work and which is allocated an establishment number according to the above-mentioned principles. A **workplace** is a unit in which employees work and which is not allocated an establishment number according to the above-mentioned principles. A **company** as a term combines establishment premises and workplaces belonging to the same employer. An **employer** is any natural person or legal entity that employs at least one employee subject to social security contributions or in marginal part-time employment. Establishment and establishment premises are synonyms; branch office is a synonym for subsidiary, district office, out-sourced office, workplace etc. if it is not an establishment.
Notes on quality	The establishment ID is only missing in a very small number of cases. These observations are notifications for the person group “205” (earnings notifications for casual workers). As establishment variables (place of work, economic activity, establishment size etc.) are merged via the establishment ID, they are missing in these observations.

6.2 Generated technical variables

6.2.1 Counter per person (spell)

Category	Description
Variable label	counter per person
Variable name	spell
Category	generated technical variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	The observation counter per person counts a person’s observations, beginning with 1. The variable is generated during the episode splitting procedure and refers to the split observations. Using the “observation counter per person” variable, it is easy to restore the original sorting order. The observations are sorted first by the start date of the split episode and then by the data source. Within employment notifications, persons subject to social insurance contributions are sorted before marginal employment notifications and higher daily wages before lower ones. One exception is one-time payments, which are sorted backwards.

6.2.2 Source of spell (quelle)

Category	Description
Variable label	Source of spell
Variable name	quelle
Category	generated technical variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical

Category	Description
Detailed description	The variable indicates the data source (see Chapter 2).

6.2.3 Year (jahr)

Category	Description
Variable label	Year
Variable name	jahr
Category	generated technical variables
Origin	BeH
Data type	numerical

Detailed description	This variable is only included in the Basis Establishment File. It indicates the year of validity of the establishment data as of the reference date of 30 June.
	This variable can be used together with the establishment number to link the Individual File and the Basis Establishment File.
	Example code with the „merge“-command in Stata 17:
	`use $orig/CILS4EU-DE-ADIAB_7523_v1.dta`
	`gen int jahr = year(begepi)`
	`sort betnr_cils4eu jahr`
	`merge m:1 betnr_cils4eu jahr using $orig/CILS4EU-DE-ADIAB_7523_v1_bhp_basis_v1.dta`

6.3 Period of validity

6.3.1 Original start date (begorig)

Category	Description
Variable label	Original start date
Variable name	begorig
Category	period of validity
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	date
Detailed description	The original start date of the observation corresponds to the original start date of the notification. This can differ from the start date of the episodes (begepi) (see also the comments on episode splitting in Section 3.2). Since the notification logic might under certain circumstances permit re-identification of the exact day of birth, the original information on the date was changed in these cases by applying the anonymisation procedure described in Section 3.1.9. **1) BeH** Because of the rules of the notification procedure, in BeH observations the starting and ending year are always identical (obligation of the employer to submit annual employment notifications). A continuous employment relationship may therefore be distributed across several notifications. **2) LHG, ASU, XASU** Certain changes lead to the creation of a new observation (see Sections 3.1.5 and 3.1.6). begorig indicates the start date of the new period. **3) MTH, XMTH**

Category	Description
	In MTH and XMTH, new data records are created when a change of provider takes place during participation in a labour market policy measure.

6.3.2 Original end date (endorig)

Category	Description
Variable label	Original end date
Variable name	endorig
Category	period of validity
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	date
Detailed description	The original end date of the observation corresponds to the original end date of the notification. This can differ from the end date of the relevant line of data, the so-called end date of the split episode (see also the comments on episode splitting in Section 3.2). Since the notification logic might under certain circumstances permit re-identification of the exact day of birth, the original information on the date was changed in these cases by applying the anonymisation procedure described in Section 3.1.9. **1) BeH** Because of the rules of the notification procedure, in BeH observations the starting and ending year are always identical (obligation of the employer to submit annual employment notifications). A continuous employment relationship may therefore be distributed across several notifications. **2) LHG, ASU, XASU** Certain changes lead to the creation of a new observation (see Sections 3.1.5 and 3.1.6). endorig indicates the end date of the new period. **3) MTH, XMTH** In MTH and XMTH, new data records are created when a change of provider takes place during participation in a labour market policy measure.

6.3.3 Episode start date (begepi)

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

6.3.4 Episode end date (endepi)

Category	Description
Variable label	Episode end date
Variable name	endepi
Category	generated period of validity

Category	Description
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	date
Detailed description	The end date of the split episode is always equal to or smaller than the end date of the original observation (see also the comments on episode splitting in Section 3.2.)

6.4 Personal information

6.4.1 Gender (frau)

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

6.4.2 Year of birth (gebjahr)

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

6.4.3 Month of birth (gebmon)

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

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

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

6.4.4 Nationality (nation)

Category	**Description**
Variable label	Nationality
Variable name	nation
Category	personal variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	The variable contains the nation codes used by the Federal Statistical Office (Statistisches Bundesamt, 2024).
Notes on quality	The variable is not filled well in the LeH before 1983.
Anonymisation	Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (nation_gr) is provided.

6.4.5 Nationality, grouped (nation_gr)

Category	**Description**
Variable label	Nationality, grouped
Variable name	nation_gr
Category	personal variables
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	The variable contains a grouped version of the nation codes used by the Federal Statistical Office (Statistisches Bundesamt, 2024). An overview of these groups can be found in the 'Overview on classifications' under the key working tools at https://fdz.iab.de/en/manuals-and-working-tools/key-working-tools/.
Notes on quality	The variable is not filled well in the LeH before 1983.

6.4.6 Marital status (famst)

Category	**Description**
Variable label	Marital status
Variable name	famst
Category	personal variables
Origin	LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	This variable describes the marital status. This variable is not filled for employment reports! **1) LeH** In the LeH, the variable has only two values (0 - not married, 1 - married). **2) LHG, ASU, XASU, MTH, XMTH**

	In the sources LHG, ASU, XASU, MTH and XMTH, a distinction is made between six values (values 11-16). The information from the different sources was not compared.
Notes on quality	**1) LeH** The quality of the information originating from the LeH is classified as poor until 2005, as it is not quite clear how the marital status was maintained. The quality is also questionable since the beginning of 2014. Between 2013 and 2017, for example, the proportion of married people drops from 37% to 14%. **2) MTH** From 2010 onwards, the share of missing values decreases. **3)LHG, XASU, XMTH** Due to deviating reporting standards in XSozial, the information from the zkT has to be classified as less valid until December 2009.

6.4.7 Number of children (kind)

Category	**Description**
Variable label	Number of children
Variable name	kind
Category	personal variables
Origin	LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	This variable has a different meaning depending on the data source. This variable is not filled for employment reports! **1) LeH** In the LeH, the variable indicates the number of children aged under 16 living in the household at the time when the application was made. It only distinguishes between whether the number of children is zero or at least one. The following values are therefore available: 0 no children 100 one or more children **2) ASU, MTH** In these sources, the value of the variable corresponds to the actual number of children under 15 living in the household. **3) LHG, XASU, XMTH** The variable reports the actual number of children aged under 15 in the benefit unit (Bedarfsgemeinschaft). In the LHG, the value is valid for the entire original period.
Notes on quality	**1) LeH** The variable is not updated when there are changes in the type of benefit or the approval of benefits, but only when a new case of benefit receipt occurs after a period of employment. This strongly restricts the quality of the data. **2) ASU, MTH** Until 30 June 2006, only up to nine children could be recorded. The value zero does not exist. For observations prior to 30 June 2006, the value zero was recoded to "missing", since it is not clear whether zero should be interpreted as "no children" or as "field not filled in". For observations after 30 June 2006, the variable is only recorded if children exist. **3) XMTH**

Category	Description
	From 2010 onwards, the share of missing values decreases.

6.4.8 Vocational training (ausbildung)

Category	Description
Variable label	Vocational training
Variable name	ausbildung
Category	personal variables
Origin	BeH, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	The variable contains the vocational training qualification. A classification of the degrees in the International Standard Classification of Education (ISCED 2011) can be found in the key working tools of the FDZ. It must be taken into account that this variable has a different meaning depending on the data source: **1) BeH** For observations obtained from the BeH, the variable contains the vocational education reported by the employers as part of the employment notification procedure. The following values exist: 1 Without vocational training 2 In-company voc. training/traineeship/external voc. training 11 University of applied sciences without further specifications 12 University without further specifications In notifications that rely on the new occupation code (see Section 5.3) it is no longer possible to identify graduates of universities of applied sciences clearly, as the new occupation code no longer has a separate category for this vocational qualification. They are assigned to category 12. **2) ASU, MTH** For these observations the vocational education completed most recently is reported. The following values exist in spells with a start date until 30 June 2006: 1 Without vocational training 2 In-company voc. training/traineeship/external voc. training 3 Technical school (voc. training) 4 Technical school (advanced voc. training) 5 University of applied sciences (FH) 6 University In 2006, the IT procedure from which the jobseeker data originate was switched from coArb to VerBIS. Many variables, such as training, were reported with different levels of differentiation in the two systems. This means that in spells from ASU and MTH which have a start date from 1 July 2006 onwards, the following differentiated categories are available: 7 Voc. training not accepted in Germany 8 University degree not accepted in Germany 9 In-company voc. training/traineeship/in-school voc. training 10 Other exams

Category	Description
	11 University of applied sciences without further specifications 12 University without further specifications 13 Doctorate 14 Bachelor (BA) 15 Bachelor (FH) 16 Bachelor (University) 17 Master (FH) 18 Master (University) 19 Diploma (BA) 20 Diploma (FH) 21 Diploma (University) 23 Undergraduate studies 24 Secondary/additional studies 25 Other (promotion-) advanced training for graduates 26 Dual study programme, integrating vocational training 27 Dual study programme, integrating practical experience 28 Bachelor 29 Master/Diploma/Magister 30 State examination, ecclesiastical examination **3) XASU, XMTH** For spells that originate from these sources, the vocational education completed most recently is reported. The following values exist: 1 Without vocational training 2 In-company voc. training/traineeship/external voc. training 3 Technical school (voc. training) 4 Technical school (advanced voc. training) 5 University of applied sciences (FH) 6 University 7 Voc. training not accepted in Germany 8 University degree not accepted in Germany For analyses that cover a longer period of time, the values can be aggregated as follows: 1, 7, 8 to 1 2, 3, 4, 9 to 2 5, 6, 10, 11-21, 23-30 to 3 The aggregated values have following value labels: 1 Without (recognised) formal vocational qualification 2 Vocational education and training (VET) 3 Academic education
Notes on quality	**1) BeH** “Changes in the vocational training status frequently occur at the same time as a change of establishment. This is because the notification data are compiled anew in the new firm. If, for example, an employee obtained a higher qualification via a part-time further training course while still working then this change of status is probably not recorded until he/she joins a new firm. It can generally be assumed

Category	Description
	that when a person is employed in a firm for a longer period, the personal data that they reported when they joined the firm is simply carried forward" (own translation of Meinken / Koch 2004, p. 63). The share of missing values increases almost continuously over time. Due to the introduction of the new occupation code in 2011, the share even temporarily strongly increased to around 51%. Since 2014, however, the proportion of missing values in the BeH has levelled off at around 40%. Missing values occur particularly frequently in the following groups: marginal part-time employees, part-time workers, foreign employees and workers from Eastern German. The reason for this is that the variable is not of particular importance as regards social security contributions (see Meinken/Koch, 2004, p. 63). The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments. For the variable "Vocational training (imputed)" (ausbildung_imp, see Section 6.4.9), a method was applied to correct missing values or inconsistent changes of the training variable. However, this variable is only filled in the source BeH and has different categories than the variable ausbildung. **2) ASU, MTH** As a result of the switchover from coArb to VerBIS it is not possible to distinguish correctly between "no completed vocational training" and "no information available" in the ASU and MTH data sources between 2006 and 2008. A missing value in this period therefore does not necessarily mean that the person has no vocational training or that there is no information available on vocational education and training, but may also mean that it was not possible to apply the relevant data generation procedure. In the source MTH, older categories partially still occur even after 1 July 2006. **2) XASU, XMTH** The degree of completeness in the XASU and the XMTH is generally low.

6.4.9 Vocational training (imputed) (ausbildung_imp)

Category	Description
Variable label	Vocational training (imputed)
Variable name	ausbildung_imp
Category	personal variables
Origin	BeH
Data type	numerical
Detailed description	The variable Vocational training (imputed) is a supplement to the variable Vocational training (ausbildung) and contains additional and harmonized information on the vocational training of employees for BeH spells. The variable thus offers a solution to the problems concerning the variable Vocational training (ausbildung) described in Section 6.4.8. The imputation procedure is described in Thomsen et al (2018), which is based on the work of Fitzenberger et al. (2006). As the variable only uses the training information from BeH notifications and because the educational categories of the old and the new occupation codes had to be harmonised for the variable Vocational training (ausbildung), the variable ausbildung_imp has other categories than the variable ausbildung.

Category	Description
	A classification of the degrees in the International Standard Classification of Education (ISCED 2011) can be found in the key working tools of the FDZ (https://fdz.iab.de/en/manuals-and-working-tools/key-working-tools/).
Notes on quality	Despite the imputation procedure, the share of missing data increases over time, reaching about 9% in 2019. The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments.

6.4.10 School leaving qualification (schule)

Category	Description
Variable label	School leaving qualification
Variable name	schule
Category	personal variables
Origin	BeH, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	This variable contains the school leaving qualification. Different values are possible depending on the source. A classification of the degrees in the International Standard Classification of Education (ISCED 2011) can be found in the key working tools of the FDZ (https://fdz.iab.de/en/manuals-and-working-tools/key-working-tools/). **1) BeH** With the switch to the new occupation code in 2011 (see Section 5.3) the possible values of the variable change. The values from the old occupation code are: 5 Grade-/lower school certificate, intermediate school or equivalent qualification 8 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college or upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling 9 Upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling With the new occupation code the values are: 1 No school leaving certificate 4 Lower secondary school certificate/ grade school certificate 6 Intermediate school leaving certificate 8 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college or upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling The differentiation options in lower secondary education were thus expanded (1, 4, 6 instead of 5 previously), while those in upper secondary education were reduced (8 and 9 now combined under 8). **2) ASU, XASU, MTH, XMTH** The following values are possible for these data sources: 1 No school leaving certificate 4 Lower secondary school certificate/ grade school certificate 6 Intermediate school leaving certificate

Category	Description
	7 Completion of education at a specialised upper secondary school/completion of higher education at a specialised college 9 Upper secondary school leaving certificate, A-level equivalent, qualification for university; 13 years of schooling They are valid at the beginning of the period of job-search or participation in a measure. In the case of people seeking an apprenticeship position, the variable may also contain the school qualification they are working towards in the XASU data source.
Notes on quality	**1) BeH** The degree of completeness in the BeH has been decreasing continuously over time and seems to have levelled off at under 2/3 in recent years. The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments. **2) ASU, XASU, MTH, XMTH** In the XASU and the XMTH, the degree of completeness increases continuously and has levelled off at over 75% since 2018 (XASU) and 2015 (XMTH), respectively. In the ASU and the MTH the degree of completeness is generally high.

6.5 Information on employment, benefit receipt and job search

6.5.1 Daily wage/daily benefit (tentgelt)

Category	Description
Variable label	Daily wage/daily benefit
Variable name	tentgelt
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH
Data type	numerical
Detailed description	**1) BeH** In BeH observations, this variable shows the employee's gross daily wage. It is calculated from the fixed-period wages reported by the employer and the duration of the (unsplit) original notification period in calendar days. The daily wage is shown in euros. Until 1998, employers in principle only reported the earnings which were subject to social security contributions. Earnings below the marginal part-time income threshold were not reported. Earnings exceeding the upper earnings limit for statutory pension insurance are only reported up to this limit. There are two upper earnings limits in the statutory pension insurance scheme. The earnings limit of the miners' pension insurance is generally higher than the earnings limit of the pension insurance for wage and salary earners. However, it is not possible to differentiate between these two insurance providers in the data. Since the inclusion of marginal part-time employees in the employment notification procedure on 1 April 1999, earnings below the marginal part-time income threshold have also been recorded; the upper earnings limit still applies as the upper ceiling. In some cases, however, the reported earnings nonetheless exceed the upper earnings limit. Generally, this can probably be attributed to the payment of annual bonuses which the employer can add to the regular earnings in the annual, employment interruption or end of employment notifications. In this

Category	Description
	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 the key working tools (https://fdz.iab.de/en/manuals-and-working-tools/key-working-tools/). A daily wage reported as 0 euros can be put down to “employment interruption notifications”. During these periods, the employment relation-ship continues to exist in legal terms, but without pay. This is the case for periods of illness after the end of continued payment of wages, for periods of maternity leave and for sabbaticals. The daily wage is shown with two decimal places. All values greater than 0 and smaller than 0.01 were rounded up to 0.01. This makes it possible to identify the above-mentioned employment interruption notifications with the condition daily wage = 0. **2) LeH** For LeH observations, the variable shows the daily benefit rate, converted into euros in each case. It must be taken into account that for observations with an original end date prior to 1 January 1998 the daily benefit rate applies to working days (i.e., including Saturdays but excluding Sundays and public holidays), while for observations with an original end date from 1 January 1998 onwards it applies to calendar days. Since 1 January 2005, a daily benefit rate reported as 0 euros can be attributed to benefit suspension periods or interruptions. Further information on these periods can be found in the variable “Reason of cancellation/notification/termination (grund)“.
Notes on quality	**1) BeH** From 2013 onwards, the number of notifications with a reason for deregistration of 54 (notification of a one-off special payment) increases sharply (see variable “Reason of cancellation/notification/termination (grund)”). It is likely that special payments which were reported with the annual declarations before 2013 are now reported separately. It is therefore advisable to add these variable one-time payments to the corresponding wages for simultaneous employment episodes within the same establishment when analysing wages over time (see variable Daily wage (incl. lump sum payment)).

6.5.2 Daily wage (incl. one-off special payment) (tentgelt_bonus)

Category	Description
Variable label	Daily wage (incl. one-off special payment)
Variable name	tentgelt_bonus
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Detailed description	The variable Daily wage (incl. one-off special payment) is a supplement to the variable Daily wage/daily benefit (tentgelt) and contains harmonised information on daily wage supplemented by one-off payments/special payments for BeH

Category	Description
	spells. Values above 97.5% of the mode of daily wages rounded to whole euro amounts for full-time male employees (per year and separately for East and West Germany) are truncated. This value is lower than the official upper earnings limit. The procedure is described in chapter 4.1 of Drechsler et al (2023).
Notes on quality	The variable is not available for all observations from the BeH source. On the one hand, episodes with notification on one-off payments/special payments (154) are not taken into account, on the other hand, Drechsler et al. (2023) clean up multiple employment for the same period in the same establishment in advance, not useing it to generate the variable. These episodes are still included in the data product presented here.

6.5.3 Daily wage (imputed) (tentgelt_imp)

Category	Description
Variable label	Daily wage (imputed)
Variable name	tentgelt_imp
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Detailed description	The variable ‘Daily wage’ (imputed) is an additional variable to the variable ‘Daily wage/daily benefit (tentgelt) and contains imputed daily wages for BeH notifications of full-time employees aged between 14 and 64 (excluding trainees) if these are censored in the notification data because they exceed the upper earnings limit. The variable is based on the variableDaily wage (incl. one-off special payment) (tentgelt_bonus). This means that the daily wages of this subpopulation that do not exceed the upper earning limits are taken from ‘tentgelt_bonus’. Details of the imputation procedure are described in Drechsler et al. (2023) and Drechsler/Ludsteck (2025).
Notes on quality	Drechsler et al. (2023, Chapter 5) warn against using the variable in analyses without further test steps. In particular, it should always be checked how high the proportion of censored wages is in the respective model and whether all regressors of the model have also found their way into the imputation model either directly or through sufficiently good proxies. It is therefore recommended that you familiarise yourself with the assumptions and limitations of the procedure before using it.

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

Category	Description
Variable label	Occupation – current/most recent (KldB 1988)
Variable name	beruf
Category	information on employment, benefit receipt and job search
Origin	BeH, ASU, XASU, MTH
Data type	numerical
Detailed description	**1) BeH** The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes applies to one employee, the employer is required to select the job title that best defines the main activity performed (see BA 2005, p. V).

Category	Description
	For this, the employer encodes the employee's job in accordance with the "Classification of Occupations. Systematic and Alphabetical Directory of Job Titles" (published by the Federal Employment Agency, Nuremberg, 1988), which contains approx. 25,000 job titles. The occupational classification consists of a 3-digit code and comprises about 330 values. Employment notifications with an end date later than 30 November 2011 are reported using the new occupation code 2010 (KldB2010) (see Section 5.3). These values are transcoded to the KldB1988 via a priority switch. This results in inaccuracies. **2) ASU, XASU, MTH** The variable contains the occupation of the last job. See 1) with regard to the occupation code.
Notes on quality	**1) BeH** There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. A similar accumulation of missing values occurs in 1991 due to reunification. The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments. **2) ASU, MTH** The variable is only filled after 2011 if the last activity was still reported according to the old occupation code. The degree of completeness decreases continuously from year to year. After 2016, it is under 10%. **3) XASU** The occupation variable is not filled for almost the entire period available.

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

Category	Description
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
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, 2021). 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 - current/most recent job (KldB 2010) (niveau)'. Employment notifications with an end date earlier than 30 November 2011 are reported using the old occupation code 1988 (KldB 1988) (see Section 5.3). These

Category	Description
	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. The KldB2010 was updated in 2020. Two new occupational subgroups (4-digit) and 14 new occupational categories (5-digit) were created. In addition, one occupational subgroup and one occupational class were moved within the classification. The BeH source contains the changes from 2020 onwards. **2) LeH, ASU, XASU, MTH** The variable contains the occupation of the last job. See 1) with regard to the occupation code. The changes of the classification 2020 were integrated retrospectively in these sources, so that they can already be found from 2006 onwards. **3) all sources** The occupational group can be assigned to the occupational sectors and occupational segments according to the classification of the Federal Employment Agency and the IAB via the main occupational group (first two digits of the key). A classification can be found in the key working tools of the FDZ (https://fdz.iab.de/en/manuals-and-working-tools/key-working-tools/).
Notes on quality	**1) BeH** There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. A similar accumulation of missing values occurs in 1991 due to reunification. The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments. **2) LeH** This variable was previously not filled in the LeH, since the occupational data in that source was of poor quality. By now, the occupation in the LeH is adopted from the job seeker history, leading to an increase in quality. **3) XASU** The variable is not filled for almost the entire period available.

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

Category	Description
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
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, 2021). 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 - current/most recent job (KldB 2010) (niveau)'. Employment notifications with an end date earlier than 30 November 2011 are reported using the old occupation code 1988 (KldB 1988) (see Section 5.3). These values are recoded to the KldB2010 by transferring the key area. As the new occupation code is considerably more detailed than the old one, this results in substantial inaccuracies. This must be taken into account when analysing the data. The KldB2010 was updated in 2020. Two new occupational subgroups (4-digit) and 14 new occupational categories (5-digit) were created. In addition, one occupational subgroup and one occupational class were moved within the classification. The BeH source contains the changes from 2020 onwards. **2) LeH, ASU, XASU, MTH** The variable contains the occupation of the last job. See 1) with regard to the occupation code. The changes of the classification 2020 were integrated retrospectively in these sources, so that they can already be found from 2006 onwards. **3) all sources** The occupational group can be assigned to the occupational sectors and occupational segments according to the classification of the Federal Employment Agency and the IAB via the main occupational group (first two digits of the key). A classification can be found in the key working tools of the FDZ (https://fdz.iab.de/en/manuals-and-working-tools/key-working-tools/). Employees working as master craftsmen can be identified, with a few exceptions in the transport, trade and personal service occupations (beruf2010_3: 514, 621, 632, 633, 713, 714, 732, 813, 831, 832, 921, 944 and 946), by combining the value 9 in the fourth position (occupational subgroup) with the value 3 in the level of requirement (niveau).
Notes on quality	**1) BeH** There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. A similar accumulation of missing values occurs in 1991 due to the reunification. The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments. **2) LeH** This variable was previously not filled in the LeH, since the occupational data in that source was of poor quality. By now, the occupation in the LeH is adopted from the job seeker history, leading to an increase in quality. **3) XASU** The variable is not filled for almost the entire period available.
Anonymisation	Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (beruf2010_3) is provided.

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

Category	Description
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
Detailed description	**1) BeH** The occupational title of the job performed by the employee during the notification period is a component of the 'employment details' submitted by the employer. If more than one job title with different classification codes apply for one employee, the employer is required to select the job title that best defines the main activity performed (see Bundesagentur für Arbeit, 2005, p. V). For this the employer encodes the employee's job in accordance with the "Classification of Occupations 2010" (Klassifikation der Berufe 2010, KldB2010, Bundesagentur für Arbeit, 2011). The occupational class consists of a 5-digit code and comprises about 1300 values. The less detailed occupational sub-group is recorded by the first four digits of the code. The skill level required for a job, which is recorded in the fifth digit of the codes in the KldB2010, is made available separately in the variable 'level of requirement' (niveau). Employment notifications with an end date earlier than 30 November 2011 are reported using the old occupation code 1988 (KldB 1988) (see Section 5.3). These values are recoded to the KldB2010 by transferring the key area. As the new occupation code is considerably more detailed than the old one, this results in substantial inaccuracies. This must be taken into account when analysing the data. The KldB 2010 was updated in 2020. Two new occupational subgroups (4-digit) and 14 new occupational categories (5-digit) were created. In addition, one occupational subgroup and one occupational class were moved within the classification. The BeH source contains the changes from 2020 onwards. **2) LeH, ASU, XASU, MTH** The variable contains the occupation of the last job. See 1) with regard to the occupation code. The changes of the classification 2020 were integrated retrospectively in these sources, so that they can already be found from 2006 onwards.
Notes on quality	**1) BeH** There is a considerable increase in the number of missing values in 2011 due to the change in the occupation code. A similar accumulation of missing values occurs in 1991 due to the reunification. The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments. **2) LeH** This variable was previously not filled in the LeH, since the occupational data in that source was of poor quality. By now, the occupation in the LeH is adopted from the job seeker history, leading to an increase in quality. **3) XASU** The variable is not filled for almost the entire period available.

6.5.8 Part-time (teilzeit)

Category	Description
Variable label	Part-time
Variable name	teilzeit
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Detailed description	The variable “Part-time” (teilzeit) distinguishes between full-time and part-time employees. The decisive factor is the ratio between the contracted hours and the usual working hours in the establishment.
Notes on quality	There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. In order to reduce this problem, the working hours were imputed at the IAB for the period in question. Further information about the procedure can be found in Ludsteck/Thomsen (2016). A similar accumulation of missing values occurs in 1991 due to the reunification. No imputation is performed here. The introduction of the new occupation code is associated with a break in the data that goes beyond the mere conversion of the key and can possibly be explained by updating effects during the changeover of the payroll accounting software at the reporting establishments. For a more detailed discussion of the problem and a possible corrective approach, see Fitzenberger/Seidlitz (2020).

6.5.9 Occupational status and working hours (stib)

Category	Description
Variable label	Occupational status and working hours
Variable name	stib
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Detailed description	The employee’s occupational status during the notification period is reported by the employer as part of the “employment details”. The variable “occupational status and working hours” distinguishes between full-time and part-time employees. The decisive factor for this differentiation is the ratio between the contracted hours and the usual working hours in the establishment. For part-time employees the variable only indicates whether their working hours exceed a certain threshold. Until 1978, this threshold was 20 hours of work per week, between 1979 and 1987 it was 15 hours per week and since 1988 it is 18 hours per week. The variable only provides information regarding the occupational status for full-time employees, distinguishing among other things between blue-collar and white-collar employees in full-time employment and apprentices. The distinction between (full-time) blue-collar and white-collar employees is solely based on the type of compulsory pension insurance (Federal Social Insurance Office for Salaried Employees – BfA – for white collar workers, and State Social Insurance Office – LVA – for blue-collar workers). Master craftsmen and foremen are only included in a separate category if they are compulsorily insured in the workers' pension insurance. The assignment of a master craftsman or foreman to the blue-collar or white-collar employees can only be made via the respective pension provider. The "employees in vocational training" category covers not only apprentices,

Category	Description
	volunteers and interns but also semi-skilled trainees, students at colleges for health occupations and participants in subsidised further vocational training, retraining and induction training. If more than one code is eligible for an employee, the employer is required to classify the job according to the activity which is predominantly performed. If this cannot be determined clearly, the code of the higher occupational status is to be entered (see BA 2005, p. VI). Owing to the introduction of the new occupation code (see Section 5.3), however, this distinction is no longer possible. The variable „stib“ is therefore only filled for notifications which date back to before the introduction of the new occupation code. However, it is possible to identify occupations that require master craftsman training using the KldB 2010. For more information, see the variable "Occupational sub-group - current/most recent (KldB 2010), 4-digit (beruf2010_4)".
Notes on quality	There is a considerable increase in the number of missing values in 1991 due to the reunification. From 2011 onwards, the variable no longer contains any valid values, as it was no longer reported.

6.5.10 Employment status (erwstat)

Category	Description
Variable label	Employment status
Variable name	erwstat
Category	information on employment, benefit receipt and job search
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	This variable takes on different values with different meanings for each data source. **1) BeH** For BeH observations, the variable ‘employment status’ corresponds to the person group recorded in the new notification procedure (DEÜV) from 1 January 1999 onwards. It indicates contribution- or benefit-related particularities of the employment relationship. If multiple codes apply to an employment notification, the smallest must be indicated by the reporting employer. The majority of these cases are employment relationships subject to social security contributions without any distinctive characteristics, which are recorded under code number 101. Accordingly, it is possible that these employment relationships are slightly overestimated. The notification procedure stipulates that changes in the employment status - e.g., when an apprentice is taken on by his/her training company after completing his/her vocational training - must be indicated by a new notification. 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

Category	Description
	whether contributions to private long-term care insurance are paid by the BA. With the introduction of unemployment benefit II in 2005, unemployment assistance, maintenance allowance and the contributions to private long-term care insurance are omitted. **3) LHG** For LHG spells, the variable shows whether the person is underage and able to work, adult and able to work, or unable to work and beyond the retirement pension limit. Since the reporting logic would make it possible to re-identify the exact date of birth in many cases, the original dates were changed by means of the anonymization procedure described in Section 3.1.9. **4) ASU, XASU** For ASU/XASU observations, the ‘employment status’ variable reports the job search status. A distinction is made between those who are "unemployed and seeking work", "not unemployed and seeking work", "seeking advice" and "not seeking work". Applicants who only want advice from the BA are considered as “seeking advice”. In addition, there are rehabilitants and, before 2008, persons aged 58 or older who are not fully available for placement. As of 1 August 2016, this may also include persons who no longer require support, but who are nevertheless still under the job center’s care. "Not seeking work" mainly subsumes persons of whom activation or placement cannot be reasonably expected according to § 10 SGB II. Similarly, persons with an incapacity to work of more than 42 days who continue to receive ALG II are listed in the system under this status. Since the reporting logic would make it possible to re-identify the exact date of birth in many cases, the original dates were changed by the anonymization procedure described in Section 3.1.9. **5) MTH, XMTH** For observations in the MTH and the XMTH, the ‘employment status’ variable indicates the measure-type category. This is the highest level in the hierarchy of the measure-type classifications of the BA.
Notes on quality	**1) BeH** 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. These remaining notifications were assigned to the group of persons "Employees subject to social security contributions without any distinctive characteristics" (101). **2) LHG** In the LHG it can be observed that there is an above-average number of 15-year-olds and to a lesser extent 16- and 17-year-olds classified as unable to work. 15- and 16-year-old benefit recipients of the authorised municipalities may therefore be under-represented, as ‘individuals who are unable to work’ are not included in the IEB. In some cases, there may be conflicting information on a person's ability to work if he or she is a member of different benefit units (BGs) at the same time. A possible reason for this is the determination of the ability to work by the job center

Category	Description
	specialist supervising the BG or by the respective responsible medical services. They can come to different results or document these at different times. A further reason may be different birthday records for simultaneous BG customers, which may affect the status of their ability to work. **3) ASU, XASU** The categories "seeking advice" and "not seeking work" have only existed since the introduction of Verbis (see Section 5.6). Due to late notifications, however, entries before 2006 can also be found. The characteristic "not seeking work" is considered to be under-recorded before 2008. In the XASU, the quality of status determination is partly limited for some institutions, especially in the first years.

6.5.11 Measure type – group (mass)

Category	Description
Variable label	Measure type – group
Variable name	mass
Category	information on employment, benefit receipt and job search
Origin	MTH, XMTH
Data type	numerical
Detailed description	The variable indicates the measure type group. This is the second highest level in the hierarchy of the measure type classifications of the BA.
Notes on quality	
Anonymisation	Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the measure type category is provided (see variable 'Employment status (erwstat)').

6.5.12 Transition zone (gleitz)

Category	Description
Variable label	Transition zone
Variable name	gleitz
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Detailed description	This variable is only available from 2003 onwards and only for BeH observations. It indicates whether the employment notification relates to employment in the low-wage sector, within the so-called transition zone (in German "Gleitzone", since July 2019 "Uebergangsbereich"), (so-called midi jobs). Employee only have 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 transition zone is based on gross pay and changes over time: 01.04.2003 to 31.12.2012: € 400.01 to € 800.00 01.01.2013 to 30.06.2019: € 450.01 to € 850.00

Category	Description
	01.07.2019 to 30.09.2022: € 450.01 to €1,300.00 01.10.2022 to 31.12.2022: € 520.01 to €1,600 From 01.01.2023: € 520.01 to € 2,000 Different lower limits exist for trainees and employees in voluntary services. In the case of several employment relationships, assignment to the transition zone is based on the total gross pay. The corresponding legislation has been in force since 1 April 2003.

6.5.13 Temporary agency work (leih)

Category	Description
Variable label	Temporary agency work
Variable name	leih
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Detailed description	The variable reports whether the person's employment is a temporary job via an employment agency. The employee is recorded with the establishment that provides the social security notifications and this is the establishment which leases out the employee (not the establishment that the employee physically works in). The variable is derived from the occupation code 2010 and is only available for notifications with an end date later than 30 November 2011.
Notes on quality	There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. By 2012, the share of missing values is down to about 3% and decreases below 1% until 2019.

6.5.14 Fixed-term contract (befrist)

Category	Description
Variable label	Fixed-term contract
Variable name	befrist
Category	information on employment, benefit receipt and job search
Origin	BeH
Data type	numerical
Detailed description	The variable reports whether the person's employment relationship is fixed-term or permanent. The variable is derived from the occupation code 2010 and is only available for notifications with an end date later than 30.11.2011.
Notes on quality	There is a considerable increase in the number of missing values in 2011 due to the change in the reporting procedure. By 2012, the share of missing values is down to about 3% and decreases under 1% until 2019.

6.5.15 Reason of cancellation/notification/termination (grund)

Category	Description
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

Category	Description
Detailed description	**1) BeH** In BeH observations, the ' Reason of cancellation/ notification/ termination ' variable indicates the reason why the employer submitted the employment notification in question to the social security agencies. However, not all of the possible reasons for submitting a notification that may occur in the context of the notification procedure are available in the IEB. For instance, the IEB only includes notifications that have information on earnings (i.e., annual, employment interruption and end of employment notifications), while initial registrations are not contained as they contain no information on earnings. However, this does not involve a loss of information, as the details from a registration are transmitted again with the following annual, employment interruption or end of employment notification. The reasons for submitting employment notifications are encoded according to the regulations of the notification procedure, which has been in effect since 1 January 1999 (in accordance with DEÜV) with a leading "1". To obtain the original coding, the value "100" must be subtracted from the coding in the variable. **2) LeH** In the LeH, this variable specifies the reason for the end of the receipt of unemployment benefits, unemployment assistance, or maintenance allowance. There is no information in the LeH about the reasons for the start of the benefit receipt, since the LeH is filled with the notifications from the employment agencies to the health insurance about completed benefit receipt durations. The variable "Residual claim" can additionally be used to distinguish between suspension and interruption periods: As a rule, the entitlement periods are not reduced in the case of interruption periods, but they are reduced in the case of suspension periods (see variable "Residual claim/planned duration (restanspruch)"). The values of this variable are only filled for the validity period provided by the technical or legal regulations. If values occur outside their validity period, they are converted to substitute values {value range 1173-1181} because their meaning cannot then be clarified with certainty. **3) LHG** The LHG observations contain the 'reason for termination of Unemployment Benefit II' and indicate the reason why current benefits have been discontinued. The 'reason for discontinuation of Unemployment Benefit II' variable refers to the individual, not to the benefit unit. If the composition of the benefit union changes the observations for all members of the benefit unit are terminated and started again on this date. The reason for termination of Unemployment Benefit II is only filled for the individual for whom a reason for termination is available. The variable is missing (.z) for all other members of the benefit union. 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 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 rules below:

Category	Description
	229 to 260, 230 to 260, 231 to 261, 232 to 260, 233 to 260, 234 to 260,2 35 to 260, 236 to 261, 237 to 266, 238 to 266, 239 to 271, 240 to 269, 242 to 265, 243 to 270, 244 to 274, 245 to 277, 246 to 267, 247 to 267, 248 to 278, 249 to 269, 250 to 275, 251 to 274, 252 to 276, 253 to 268, 254 to 278 **5) XASU** In the case of XASU observations, the variable contains the deregistration reason. In the XASU, overlap-free periods relating to a customer are shown. When certain characteristics change, a new episode is created for the XASU, of which the following are part of CILS4EU-DE-ADIAB: • Change of job search status • Change of availability • Change of SGB II provider • Change of place of residence If the episode reported by the municipal institutions responsible for implementing SGB II is artificially split for other reasons (see variable “Episode start date (begepi)”), "Generated because of data splitting" is stated as the termination reason.
Notes on quality	**1) BeH** From 2013 onwards, the number of notifications with a reason for deregistration of 54 (notification of a one-off special payment) increases sharply. As long as an employment relationship exists, special payments that are paid out by March of the following year can be included into the usual notifications (mostly annual notifications) for the previous year. A separate notification with a reason for deregistration of 54 is then not required. Until 2012, the annual notifications could be submitted until mid-April; since 2013, they must now be submitted by mid-February at the latest. Special payments made in February and March must now be reported separately. **2) LHG** The degree of completeness for the reason for notification in the LHG data sources is very small (< 20%) across all years.

6.5.16 Employment status prior to job search (estatvor)

Category	Description
Variable label	Employment status prior to job search
Variable name	estatvor
Category	information on employment, benefit receipt and job search
Origin	ASU, XASU
Data type	numerical
Detailed description	For ASU and XASU observations, this variable shows the employment status prior to the job search activity. 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 CILS4EU-DE-ADIAB. This information does not refer to the start date of the episode, but to the start of the original time period.
Notes on quality	**1) ASU**

Category	Description
	According with the change from coArb to VerBIS in 2005/2006 (see Section 5.6), there is a clear break in the values of this variable, as consistent mapping is not possible.

6.5.17 Employment status after job search (estatnach)

Category	Description
Variable label	Employment status after job search
Variable name	estatnach
Category	information on employment, benefit receipt and job search
Origin	ASU
Data type	numerical
Detailed description	The variable contains the person's status after leaving unemployment. Longer periods of illness can be identified via this variable. The values are classified as follows: • 1000s: measure (assisted employment) • 2000s: non-assisted employment • 3000s: training etc. • 4000s: self-employment • 5000s: exclusion/suspension • 6000s: other
Notes on quality	**1) ASU** According with the change from coArb to VerBIS in 2005/2006 (see Section 5.6), there is a clear break in the values of this variable, as consistent mapping is not possible. However, the relative frequencies also vary over time depending on the values of the variable. Particular caution is therefore required in analyses.

6.5.18 Integration forecast (ipo)

Category	Description
Variable label	Integration forecast
Variable name	ipo
Category	information on employment, benefit receipt and job search
Origin	ASU, MTH
Data type	numerical
Detailed description	The variable reports the integration forecasts assessed by the placement, counseling and integration specialist (or the pilot in SGB III) during the initial contact. The integration forecast is the result of the assessment of the client's proximity to the market. "Close to market" applies if integration is expected within six months, "not close to market" if integration is only expected to succeed after more than six months. Identifiable needs for action that can be remedied within six months do not rule out market proximity. The integration prognosis "integrated, but in need" is only valid for applicants from the SGB II. If there is only a desire for placement in training, no integration forecast is to be determined.

Category	Description
	This variable will be available from the end of 2016 and is the successor to the client profile (profil) variable. For the period before that, the information from the client profile is recoded into the integration forecast as follows: The old values 2 (market profile), 3 (empowerment profile) and 4 (incentive profile) become the new value 1 (Close to market), the old values 5 development profile), 6 (stability profile) and 7 (benefit profile) become the new value 2 (Not close to market), the old value 8 (assignment unnecessary) becomes 3 (Assignment unnecessary) and 9 (integrated, but in need) becomes 4 (Integrated, but in need) and 1 (still undetermined) becomes .z (No entry).
Notes on quality	The variable was introduced in 2016 but was recoded back to 2006 using other variables. The quality has been assessed as reliable by the BA statistics department since 2010.

6.5.19 Reason for end of previous employment (art_kuend)

Category	Description
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
Detailed description	This variable describes how the last employment or training relationship was terminated before a period of job search. It can therefore be used to identify job-to-job placements.
Notes on quality	The share of missing values is particularly high in this variable, averaging around 50%.

6.5.20 Working hours of job application (arbzeit)

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

6.5.21 Residual claim/planned duration (restanspruch)

Category	Description
Variable label	Residual claim/planned duration
Variable name	restanspruch
Category	information on employment, benefit receipt and job search
Origin	LeH, MTH
Data type	numerical
Detailed description	The variable has a different meaning depending on the data source. **1) LeH**

Category	Description
	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 end date of the split episode, but to the end of the original time period. **2) MTH** The variable contains the planned duration of the measure.
Notes on quality	**1) LeH** Over the course of a benefit receipt biography, a systematic development of the remaining entitlement is to be expected. This means that the remaining entitlement at the end of a benefit period minus the duration of the subsequent period should give the remaining entitlement after the subsequent period as long as no claims have arisen or expired in the meantime. However, there may be unexpected increases or decreases in the remaining entitlement. These are presumably mainly caused by corrections during the processing of benefit cases. Such corrections occur, for example, if a claim was not correctly determined at the start of the benefit case due to incomplete information or if a remaining claim that has not yet expired was not taken into account. However, these corrections are usually only administered for the current record, i.e. there is no correction of the complete case. In addition, in individual cases the information on entitlement durations may exceed the individual upper limit according to the statutory requirements. **2) MTH** In very few cases, the planned duration of the measure contains negative values or values greater than 5 years. These cases are recoded to missing (.z).

6.5.22 Type of provider (traeger)

Category	Description
Variable label	Type of provider
Variable name	traeger
Category	information on employment, benefit receipt and job search
Origin	LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Detailed description	The variable contains the type of institution responsible for implementing Unemployment Benefit II (LHG), the type of institution providing the measure (MTH/XMTH) or the institution responsible for managing the applicant profile (ASU/XASU). The variable contains not only the "Bundesagentur für Arbeit (BA)" as the type of institution responsible for implementing Social Code Book III (SGB III), but also three types of institution responsible for implementing Social Code Book II (SGB II). For further information see Section 2.5. The type of institutions responsible for implementing SGB II may change over time (e.g., from a

Category	Description
	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.

6.5.23 Start date of unemployment (alo_beg)

Category	Description
Variable label	Start date of unemployment
Variable name	alo_beg
Category	information on employment, benefit receipt and job search
Origin	ASU, XASU
Data type	numerical
Detailed description	The variable is filled for all sources and considers unemployment episodes (ALO) from ASU- and XASU for the calculation. The variable reports the start date of an uninterrupted sequence of periods of unemployment and is valid at the beginning of the observation (begepi). The following gaps do not result in an interruption of the period of unemployment: • any gap lasting seven days or less • periods of illness lasting up to 42 days

6.5.24 Duration of unemployment (alo_dau)

Category	Description
Variable label	Duration of unemployment
Variable name	alo_dau
Category	information on employment, benefit receipt and job search
Origin	ASU, XASU
Data type	numerical
Detailed description	The variable is filled for all sources and considers unemployment episodes (ALO) from ASU- and XASU for the calculation. The variable reports the duration (in days) of an uninterrupted sequence of periods of unemployment and is valid at the beginning of the observation (begepi). The following gaps do not result in an interruption of the period of unemployment: • any gap lasting seven days or less • periods of illness lasting up to 42 days However, when calculating the duration these gaps are not considered. Therefore, the duration in this variable may differ from the result of the duration calculation based on the "Start date of unemployment". Prior to 1997, the value “0” does not mean that the individual was not unemployed, as the ASU/XASU sources are not available here.

6.6 Location data

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

Category	Description
Variable label	Place of residence - district (Kreis)
Variable name	wo_kreis
Category	location data
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Hierarchy	federal state | district
Detailed description	In BeH and LeH observations, the place of residence at district level is only available for the years from 1999 onwards. The variable indicates the district (urban district or rural district) in which the social security contributor lives. The first two digits of the 5-digit district code (Kreisschlüssel) show the code for the federal state (Bundesland, NUTS 1), and positions 1-3 indicate the regional authority (Regierungsbezirk, NUTS 2). Federal states without a regional authority have a 0 in the third position. In the NUTS classification (Nomenclature des unités territoriales statistiques) of the European Union, districts correspond to the level NUTS 3. For the LHG and XASU sources, the place of residence applies to the whole period of the original observation. For the BeH, ASU, LeH, MTH and XMTH, the variable contains the place of residence at the beginning of the original period of time. This means that the longer the spell lasts, the higher the risk that the place of residence will become obsolete and that the information given for later dates will be incorrect. In order to guarantee consistent regional allocations across the entire observation period, the information on the district was recoded with reference to the territorial allocation of 31 December 2023 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 2023. As the district boundaries have changed over time, cases would occur in which the district code changes without the individual concerned having relocated if the territorial allocations of the districts were not updated.
Notes on quality	There are inaccuracies in the information provided in the BeH for some employees with regard to where they live. The reporting requirement does not clarify which residence - main or secondary residence with predominant residence - is to be reported by the employer. In the employment history, this can lead to the impression of "long-distance commuters" between the registered main residence and the place of work, even though the person is in fact employed at the secondary residence, i.e. does not actually commute. In the year 2015, the BeH data show a reduction in the change of residence of approx. 10-15% compared to the usual level of the surrounding years, with regional differences occurring. It has not yet been possible to determine the exact reason for this deviation.
Anonymisation	Due to its particular sensitivity with regard to data privacy, this sensitive variable is only made available on application and only in well-founded cases. By default, only the coarsened variable (wo_bula) is provided.

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

Category	Description
Variable label	Place of residence - federal state (Bundesland)
Variable name	wo_bula
Category	location data
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Hierarchy	federal state | district
Detailed description	This variable is an aggregation of the "district" variable to the 16 German federal states. In BeH and LeH observations, the federal state of the place of residence is only available for the years from 1999 onwards. Further information on the district of the place of residence can be found under Place of residence - district (Kreis) (wo_kreis)

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

Category	Description
Variable label	Place of residence - employment agency (Arbeitsagentur)
Variable name	wo_aa
Category	location data
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Hierarchy	regional directorate | employment agency
Detailed description	From 1999 onwards, this variable contains the agency district of the employment agency that is responsible for the employee's / BA client's place of residence. This information is determined from the residence address. For the LHG and XASU data sources, the place of residence is valid for the period of the original observation. In the case of the BeH, LeH, ASU, MTH and XMTH, the variable contains the place of residence at the start of the period of unemployment or job search. Accordingly, the longer the spell lasts, the higher the risk that the place of residence will become obsolete and that the information given for later dates will be incorrect. For the BeH and LeH, it is available from 1999 onwards. In order to guarantee consistent regional allocations across the entire observation period, the information on the agency district was recoded to the territorial allocation of 31 December 2023 for all data sources, i.e. in all calendar years, a place of residence is assigned to an agency district in accordance with the boundaries that the agency district had on 31 December 2023. As the boundaries of the agency district have changed over time, cases would occur in which a person's employment agency area changes without him/her having relocated if the territorial allocations were not updated. Berlin constitutes a problematic case with regard to updating territorial allocations, however: The boundaries of Berlin's employment agency areas have been changed repeatedly over the years, which could not be fully corrected even by recoding the territorial allocations. Berlin should preferably be analysed as a unit and not separately for East and West, because since the territorial reform of the employment offices in Berlin on 1 July 1997 their clear allocation to East and West is no longer possible. The BA statistics uniformly assigns Berlin to the East. Prior to 1999 the variable contains the employment agency that last processed the case of benefit receipt. It is not determined from the residence address and is only available for the LeH data source. Moreover, it is not possible to update the

Category	Description
	territorial allocations. This must be taken into account when conducting calculations over time (e.g., regional unemployment figures).
Anonymisation	Owing to its particular sensitivity with regard to data privacy, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the area of the regional directorate in which the social security contributor's place of residence is located is shown.

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

Category	Description
Variable label	Place of residence - regional directorate (Regionaldirektion)
Variable name	wo_rd
Category	location data
Origin	BeH, LeH, LHG, ASU, XASU, MTH, XMTH
Data type	numerical
Hierarchy	regional directorate | employment agency
Detailed description	This variable is an aggregation of the variable 'place of residence: employment agency' at the level of the regional directorates. Further information can be found there.

6.7 Establishment variables

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

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

6.7.2 Classification of economic activities 08, groups (w08_3)

Category	Description

Variable label	Classification of economic activities 08, groups
Variable name	w08_3
Category	establishment variables
Origin	BHP
Data type	numerical
Hierarchy	section (1-digit code) | division (2-digit code) | group (3-digit code) | class (4-digit code) | sub-class (5 digit code) of economic activity
Detailed description	This variable indicates the economic activity as a 3-digit code in accordance with the WZ08 classification and is available for the years since 2008. WZ08 stands for the "Classification of Economic Activities, Edition 2008" ("Klassifikation der Wirtschaftszweige Ausgabe 2008") of the Federal Statistical Office (eds.). The WZ08 is based on the Statistical Classification of Economic Activities in the European Community NACE Rev. 2. Each establishment is only assigned one code. If an establishment is active in different economic sectors, the main economic activity should be reflected.

6.7.3 w73_3 completed by extrapolation/imputation (w73_3_gen)

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

6.7.4 Type of imputation w73_3 (group_w73_3)

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

6.7.5 w93_3 completed by extrapolation/imputation (w93_3_gen)

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

6.7.6 Type of imputation w93_3 (group_w93_3)

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

6.7.7 w08_3 completed by extrapolation/imputation (w08_3_gen)

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

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

6.7.8 Type of imputation w08_3 (group_w08_3)

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

6.7.9 Year of first appearance (grd_jahr)

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

6.7.10 Year of last appearance (lzt_jahr)

Category	**Description**
Variable label	Year of last appearance
Variable name	lzt_jahr
Category	establishment variables
Origin	BHP
Data type	numerical

Category	Description
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 2023, it could indicate the closure of the establishment. However, other possible reasons for this are an “arbitrary change of the establishment number following a change of owner or a change in the legal form of the establishment”, the “outsourcing of parts of the firm under a new establishment number” or other administrative changes (see Bender et. al. 1996 or Bundesagentur für Arbeit 2007, pp. 9-11). An establishment does not necessarily have to be included in the BHP in the year of its last appearance, since only the key date June 30 is relevant for this inclusion. If the establishment has no employees on June 30 of its year of closure, it consequently does not appear in the BHP in that year.

6.7.11 Total number of employees (az_ges)

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

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

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

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

Category	Description
Variable label	No. marginal part-time workers
Variable name	az_gf
Category	generated establishment variables
Origin	BHP
Data type	numerical

Detailed description	The number of employees in marginal part-time employment is generated using the person group code – values 109 and 209. This variable has only been contained in the dataset since 1999 as it has only been included in the social security notification procedure since that year.

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

Category	**Description**
Variable label	Mean imp. wage all full-time employees
Variable name	te_imp_mw
Category	generated establishment variables
Origin	BHP
Data type	numerical
Detailed description	This variable contains the mean imputed gross daily wage of the full-time employees in an establishment. It does not include the wages of marginally part-time staff, apprentices or individuals participating in partial retirement schemes. The values are reported in euros for all years. According to the social security notification regulations, employers must indicate the employee's gross wage subject to social security contributions for a certain period of time (fixed period wage). Until the end of 1998, employers had to report the gross wage subject to social security contributions only. So only wages above the marginal part-time income threshold and below the contribution assessment ceiling were recorded. Since 1999, wages below the marginal part-time income threshold have also been recorded as part of the new notification procedure. Gross wages above the contribution assessment ceiling, however, are still cut. In order to calculate the gross daily wage, the fixed period wage is divided by the number of calendar days in the period. To calculate the mean, these censored wages were imputed (see Section 3.1.3.3 in Ganzer et al. 2020). These data were then aggregated at establishment level. The values are rounded to two decimal places. However, due to the "storage type" in Stata, additional decimal places are displayed that are not correct.

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

Category	**Description**
Variable label	Place of work - district (Kreis)
Variable name	ao_kreis
Category	location data
Origin	BHP
Data type	numerical
Hierarchy	federal state | district
Detailed description	The variable indicates the district (urban district or rural district) in which the employee's establishment is located. The first two digits of the 5-digit district code (Kreisschlüssel) show the code for the federal state (Bundesland, NUTS 1), positions 1-3 indicate the regional authority (Regierungsbezirk, NUTS 2). Federal states without a regional authority have a 0 in the third position. In the NUTS classification (Nomenclature des unités territoriales statistiques) of the European Union, districts correspond to the level NUTS 3. In order to guarantee consistent regional allocations across the entire observation period, the information on the district was recoded to the territorial allocation of 31 December 2023, 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 2023. As the district boundaries have changed over time, cases would occur in which the district code of the location of the establishment would change without the establishment concerned having relocated, if the territorial allocations were not updated.
Anonymisation	Owing to its particular sensitivity with regard to data protection legislation, this variable is only made available in non-aggregated form on application and only in well-founded cases. Otherwise, only the federal state (ao_bula) is shown as regional information.

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

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

6.8 Linkage variables

6.8.1 Availability of administrative personal data (match_admin)

Category	**Description**
Variable label	availability of administrative personal data
Variable name	match_admin
Category	Linkage-relevant variable
Origin	generated
Data type	numerical
Detailed description	This variable is only included in the linkage file. For every respondent in the CILS4EU-DE 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)

Category	**Description**
Variable label	Type of match
Variable name	match_typ
Category	Linkage-relevant variable
Origin	generated
Data type	numerical

Category	Description
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.

7 References

Antoni, Manfred/ Schnell, Rainer (2019). The past, present and future of the German Record Linkage Center (GRLC). In: Jahrbücher für Nationalökonomie und Statistik. Journal of Economics and Statistics, De Gruyter, Jg. 239 (2), S. 1-13.

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

Bertat, Thomas / Dundler, Agnes / Grimm, Christopher / Kiewitt, Jochen / Schomaker, Christine / Schridde, Dr. Henning / Zemann, Dr. Christian (2013): Neue Erhebungsinhalte 'Arbeitszeit', 'ausgeübte Tätigkeit' sowie 'Schul- und Berufsabschluss' in der Beschäftigungsstatistik. Methodenbericht, Bundesagentur für Arbeit – Statistik, URL: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Methodik-Qualitaet/Methodenberichte/Beschaeftigungsstatistik/Generische-Publikationen/MB-Arbeitszeit-ausgeuebte-Taetigkeit-Schul-und-Berufsabschluss.pdf?__blob=publicationFile&v=4 (accessed: 10 March 2025).

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

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

Bundesagentur für Arbeit (2009): Klassifikation der Wirtschaftszweige 1973, Nuremberg. URL: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Wirtschaftszweige/Klassifikation-der-Wirtschaftszweige-1973-2003/Generische-Publikationen/WS-1973-Excel.xlsx?__blob=publicationFile&v=6 (accessed: 10 March 2025)

Bundesagentur für Arbeit (2021): Klassifikation der Berufe 2010 – überarbeitete Fassung 2020 Band 1: Systematischer und alphabetischer Teil mit Erläuterungen. Nürnberg. URL: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Printausgabe-KldB-2010-Fassung2020/Generische-Publikationen/KldB2010-PDF-Version-Band1-Fassung2020.pdf?__blob=publicationFile&v=21). accessed: 10 March 2025).

Bundesanstalt für Arbeit (1988): Klassifikation der Berufe – Systematisches und Alphabetisches Verzeichnis der Berufsbenennung, Nuremberg. URL https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB1975-1992/Generische-Publikationen/KldB1988-Systematisches-und-alphabetisches-Verzeichnis-der-Berufsbenennungen-PDF.pdf?__blob=publicationFile&v=12 (accessed: 10 March 2025)

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

CILS4EU (2016): Children of Immigrants Longitudinal Survey in Four European Countries. Technical Report. Wave 1 – 2010/2011, v1.2.0. Mannheim.

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

Drechsler, Jörg/ Ludsteck, Johannes (2025): Imputation strategies for rightcensored wages in longitudinal datasets. J Labour Market Res 59, 19. DOI: https://doi.org/10.1186/s12651-025-00410-4 (accessed: 26 September 2025)

Drechsler, Jörg/ Ludsteck, Johannes/ Moczall, Andreas (2023): Imputation der rechtszensierten Tagesentgelte für die BeH. FDZ-Methodenreport 05/2023 (de), Nürnberg, DOI:10.5164/IAB.FDZM.2305.de.v1

Eberle, Johanna/ Jacobebbinghaus, Peter/ Ludsteck, Johannes/ Witter, Julia (2011): Generation of time-consistent industry codes in the face of classification changes * Simple heuristic based on the Establishment History Panel (BHP). FDZ Methodenreport, 05/2011 (en), Nuremberg.

Fitzenberger, Bernd/ Osikominu, Aderonke/ Völter, Robert (2006): Imputation rules to improve the education variable in the IAB employment subsample. In: Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften, vol. 126, issue 3, pp. 405-436.

Fitzenberger, Bernd/ Seidlitz, Arnim (2020): The 2011 break in the part-time indicator and the evolution of wage inequality in Germany. Journal for Labour Market Research 01/2020.

Ganzer, Andreas/ Schmucker, Alexandra/ Wolter, Stefanie (2024): Establishment History Panel 1975–2023. FDZ-Datenreport 09/2024 (en), Nürnberg. DOI: 10.5164/IAB.FDZD.2409.en.v1

Lochner, Benjamin; Wolter, Stefanie (2025): AKM effects for German labour market data 1985-2023. FDZ-Methodenreport 03/2025 (en), Nürnberg. DOI: 10.5164/IAB.FDZM.2503.en.v1

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

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

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

Schild, Christopher_Johannes/ Antoni, Manfred. (2014). Linking survey data with administrative social security data–The project „Interactions between Capabilities in Work and Private Life“. German RLC Working Paper Nr. 2 WP-GRLC-2014-02.

Schiel, Stefan/ Sandbrink, Katharina/ Aust, Folkert/ Chudziak, Nina/ Kleudgen, Martin (2016): Methodenbericht. CILS4EU-Projekt: CAPI-Befragung von jungen Erwachsenen in Deutschland. Bonn: infas Institut für angewandte Sozialwissenschaft GmbH.

Schmucker, A., Seth, S., & vom Berge, P. (2025): Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975 - 2023. FDZ Datenreport, 02/2025 (de), Nürnberg. DOI:10.5164/IAB.FDZD.2502.de.v1

Schnell, Rainer/ Bachteler, Tobias/ Bender, Stefan (2004). A toolbox for record linkage. Austrian Journal of Statistics, Jg. 33(1-2), S. 125-133.

Statistisches Bundesamt (2002): Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ 93), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-

Wirtschaftsklassifikationen/Downloads/klassifikation-wz-1993.html (accessed: 10 March 2025).

Statistisches Bundesamt (2003): Klassifikation der Wirtschaftszweige, Ausgabe 2003 (WZ 2003), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/Downloads/klassifikation-wz-2003-erlaeuterung.pdf?__blob=publicationFile (accessed: 10 March 2025).

Statistisches Bundesamt (2008): Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Gueter-Wirtschaftsklassifikationen/klassifikation-wz-2008.html (accessed: 10 March 2025)

Statistisches Bundesamt (2024): Staats- und Gebietssystematik, Wiesbaden. URL: https://www.destatis.de/DE/Methoden/Klassifikationen/Staat-Gebietsystematik/Staatsangehoerigkeitsgebietsschluessel_pdf.pdf?__blob=publicationFile&v=5 (accessed: 10 March 2025)

Statistik der Bundesagentur für Arbeit (2021): Grundlagen: Methodenbericht – Einführung der „Klassifikation der Berufe 2010 – über-arbeitete Fassung 2020“ in die Arbeitsmarktstatistiken, Nürnberg. URL: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Methodik-Qualitaet/Methodenberichte/Uebergreifend/Generische-Publikationen/Methodenbericht-Klassifikation-Berufe-ueberarbeitete-Fassung.pdf?__blob=publicationFile&v=4 (accessed: 10 March 2025).

Stüber, Heiko/ Dauth, Wolfgang/ Eppelsheimer, Johann (2023): A guide to preparing the sample of integrated labour market biographies (SIAB, version 7519 v1) for scientific analysis. Journal for Labour Market Research, Volume 57, Article number: 7. https://doi.org/10.1186/s12651-023-00335-w.

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

Wermter, Winfried/ Cramer, Ulrich (1988): Wie hoch war der Beschäftigtenanstieg seit 1983? – Ein Diskussionsbeitrag aus der Sicht der Beschäftigtenstatistik der Bundesanstalt für Arbeit. In: Mitteilungen aus der Arbeitsmarkt – und Berufsforschung 4/88, pp. 468-482.

8 Appendix

8.1 Labels, frequency tables and missing values

The online appendix to the data report provides detailed overviews of the individual variables in the data product. It contains label lists of the individual categories, counts and overviews of the proportion of missing values broken down by data source. You can find the appendix in separate files in German and English at: https://fdz.iab.de/

- https://doku.iab.de/fdz/reporte/2025/DR_05-25_frequencies_labels_de.zip
- https://doku.iab.de/fdz/reporte/2025/DR_05-25_frequencies_labels_en.zip

8.2 List of abbreviations

AA	Agentur für Arbeit / Arbeitsamt	employment agency / employment office
ALG	Arbeitslosengeld	unemployment benefit
ARGE	Arbeitsgemeinschaft	cooperation of employment agencies and municipalities
ASU	Arbeitsuchendenhistorik	Jobseeker History
A2LL	Arbeitslosengeld II – Leistungen zum Lebensunterhalt	unemployment benefit II - benefits to secure a livelihood
BA	Bundesagentur für Arbeit	Federal Employment Agency
BeH	Beschäftigtenhistorik	Employee History
BfA	Bundesversicherungsanstalt für Angestellte	Federal Social Insurance Office for Salaried Employees
BG	Bedarfsgemeinschaft	Benefit unit
BHP	Betriebs-Historik-Panel	Establishment History Panel
BMAS	Bundesministerium für Arbeit und Soziales	Federal Ministry of Labour and Social Affairs
CILS4EU-DE	Children of Immigrants Longitudinal Survey in Four European Countries- - Germany	Children of Immigrants Longitudinal Survey in Four European Countries - Germany
coArb	Computerunterstützte Arbeitsvermittlung (operatives Verfahren zur Verwaltung der Vermittlung (Altverfahren))	computer-aided job placement (procedure for the administration of job placements – old procedure)
DEÜV	Verordnung über die Erfassung und Übermittlung von Daten für die Träger der Sozialversicherung – Datenerfassungs- und –übermittlungsverordnung	Data Collection and Transmission Regulation - regulation on the collection and transmission of data for the social security agencies
DEVO	Zweite VO über die Erfassung von Daten für die Träger der Sozialversicherung und für die BA – Datenerfassungs-Verordnung –	Data Collection Regulation - second regulation on the collection of data for the social security agencies and for the Federal Employment Agency
DIM	Daten- und IT-Management	Data and IT Management
DÜVO	Zweite VO über die Datenübermittlung auf maschinell verwertbaren	Data Transmission Regulation - second regulation on the transfer of data on

	Datenträgern im Bereich der Sozialversicherung und der BA – Datenübermittlungs-Verordnung –	machine-readable data media in the field of social security and the BA
EDV	Elektronische Datenverarbeitung	Electronic data processing
FDZ	Forschungsdatenzentrum	Research Data Centre
FELEG	Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit	Act on the Support in Case of Termination of Farming Activities
gAw	Träger mit getrennter Aufgabenwahrnehmung	Municipalities exercising their duties separately
gE	Gemeinsame Einrichtung	Joint facility
gT	Getrennte Trägerschaft	Separated responsibilities
IAB	Institut für Arbeitsmarkt- und Berufsforschung	Institute for Employment Research
IEB	Integrierte Erwerbsbiographien	Integrated Employment Biographies
ISIC	International Standard Industrial Classification of All Economic Activities	International Standard Industrial Classification of All Economic Activities
KldB	Klassifikation der Berufe	Classification of Occupations
LeH	Leistungsempfängerhistorik	Benefit Recipient History
LHG	Leistungshistorik Grundsicherung	Unemployment Benefit II Recipient History
LVA	Landesversicherungsanstalt	Land Social Insurance Office
MTH	Maßnahmeteilnahmehistorik	Participants-in-Measures History File
MZES	Mannheimer Zentrum für Europäische Sozialforschung	Mannheim Centre for European Social Research
NACE	Nomenclature génerale des activités économiques dans les communautés européennes	Nomenclature génerale des activités économiques dans les communautés européennes
NUTS	Nomenclature des unités territoriales statistiques	Nomenclature des unités territoriales statistiques
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	Arbeitsuchendenhistorik aus XSozial-BA-SGB II	Jobseeker History from XSozial-BA-SGB II
XMTH	Maßnahmeteilnahmehistorik aus XSozial-BA-SGB II	Participants-in-Measures History File from XSozial-BA-SGB II
zkT	Zugelassene kommunale Träger	Authorised municipalities

Imprint

FDZ-Datenreport 05|2025 EN

Date of publication
30 October 2025

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

Download
https://doku.iab.de/fdz/reporte/2025/DR_05-25_EN.pdf

Documentation version
CILS4EU-DE-ADIAB7523_EN_v1_dok1, DOI: 10.5164/IAB.FDZD.2505.en.v1

Dataset version
CILS4EU-DE-ADIAB 7523 v1, DOI: 10.5164/IAB.CILS4EU-DE-ADIAB7523.de.en.v1

All publications in the series "FDZ-Datenreport" can be downloaded from
https://fdz.iab.de/en/research/publications/fdz-datenreport-series/

Website
https://fdz.iab.de

Corresponding author
Alexandra Schmucker
Phone: +49 911 179-1752
Email: Alexandra.Schmucker@iab.de