

# FDZ *Datenreport*

---

Documentation of labour market data

No. 03/2009 (EN)

## Integrated Employment Biographies Sample

IEBS

*Dirk Oberschachtsiek, Patrycja Scioch,*

*Christian Seysen, Jörg Heining*

**Handbook  
for the IEBS in the 2008 version**



**Bundesagentur für Arbeit**

## Contents

<b>1</b>	<b>Introduction.....</b>	<b>5</b>
<b>2</b>	<b>Outline of the dataset.....</b>	<b>7</b>
2.1	Overview .....	8
2.2	Volume structure and degree of completeness over time .....	9
2.3	Differences between the IEBS and the IABS and the BA Employment Panel .....	11
<b>3</b>	<b>Data preparation.....</b>	<b>11</b>
3.1	Data sources .....	11
3.1.1	Employee History of the IAB (BeH) .....	11
3.1.2	Benefit Recipient History of the IAB (LeH) .....	12
3.1.3	Participation-in-Measures History File (MTH).....	13
3.1.4	Job-search status from the Applicant Pool Data (BewA) .....	14
3.2	Linking the data .....	14
3.3	Corrections and filters .....	16
3.3.1	BeH .....	16
3.3.2	LeH.....	17
3.3.3	MTH .....	18
3.3.4	BewA.....	19
3.3.5	Aggregate file.....	20
3.4	Legally inadmissible parallel states.....	20
3.5	Data quality and problems .....	21
3.5.1	Employee History (BeH) .....	22
3.5.2	Benefit Recipient History (LeH) .....	22
3.5.3	Participants-in-Measures History File (MTH).....	23
3.5.4	Applicant Pool Data (ASU / BewA).....	23
3.6	Episode splitting .....	25
3.7	Anonymisation of data and sensitive variables.....	27
3.8	Test data and comments regarding data evaluations.....	27
<b>4</b>	<b>Description of variables.....</b>	<b>28</b>
4.1	General part .....	28
4.1.1	Overview of variables.....	28
4.1.2	Missing values .....	31
4.2	Identification numbers .....	31
4.2.1	Individual ID .....	31
4.2.2	Data record number .....	32
4.2.3	Establishment number .....	32
4.3	Start and end dates of observation .....	32
4.3.1	Original start date of observation.....	32
4.3.2	Original end date of observation.....	32
4.3.3	Start date of split episode .....	33
4.3.4	Start and end dates.....	33
4.4	Generated technical variables .....	34
4.4.1	Source of the information.....	34
4.4.2	Source combination .....	34

4.4.3	Observation counter per person .....	35
4.4.4	Number of observations per person .....	35
4.4.5	Observation counter per episode.....	36
4.4.6	Number of simultaneous observations per episode .....	36
4.4.7	Observation counter per episode and source.....	36
4.4.8	Number of simultaneous observations per episode and source .....	37
4.5	<i>Personal status before, during and after the current observation.....</i>	<i>38</i>
4.5.1	Employment status prior to job-search .....	38
4.5.2	Employment status: person group, type of benefit, type of measure, job-search status .....	38
4.5.3	Reason for end of observation.....	39
4.5.4	Status after exit .....	40
4.6	<i>Personal variables.....</i>	<i>41</i>
4.6.1	Year of birth .....	41
4.6.2	Gender .....	42
4.6.3	Nationality .....	42
4.6.4	Severe disability status .....	42
4.6.5	School-leaving qualification .....	43
4.6.6	School education and vocational training .....	43
4.6.7	Skills level .....	44
4.6.8	Marital status.....	45
4.6.9	Number of children in the household .....	45
4.6.10	Health impediments .....	46
4.7	<i>Data on the employment relationship and job-search activity .....</i>	<i>47</i>
4.7.1	Occupational status and working hours.....	47
4.7.2	Occupation .....	48
4.7.3	Daily wage / daily benefit rate.....	48
4.7.4	Transition zone .....	49
4.7.5	Classification of economic activities (73, 93, 03).....	49
4.7.6	Willingness to seek employment throughout Germany .....	50
4.7.7	BA client group.....	51
4.7.8	Type of termination .....	51
4.7.9	Desired working hours of the job sought .....	52
4.7.10	Duration of remaining entitlement to unemployment benefit .....	52
4.7.11	Planned start date.....	52
4.7.12	Planned end date.....	53
4.8	<i>Data on the employing establishment.....</i>	<i>53</i>
4.8.1	Number of regular employees .....	53
4.8.2	Year when establishment was founded.....	54
4.9	<i>Location data.....</i>	<i>55</i>
4.9.1	Place of work: federal state (Bundesland), district (Kreis).....	55
4.9.2	Place of residence: federal state (Bundesland), district (Kreis).....	55
4.9.3	Place of work: regional directorate (employment agency area) .....	56
4.9.4	Place of residence: regional directorate (employment agency area) .....	57
4.9.5	Place of residence: employment agency region type .....	58
4.9.6	Place of residence abroad .....	58
<b>5</b>	<b>References .....</b>	<b>59</b>
<b>6</b>	<b>Appendix .....</b>	<b>60</b>
6.1	<i>List of abbreviations .....</i>	<i>60</i>
6.2	<i>Alphabetical list of variables.....</i>	<i>62</i>

**Data availability**

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

**Acknowledgements**

We would like to thank our colleagues in the ITM division of the Institute for Employment Research (IAB) for their extensive preliminary work and their assistance.

The authors of the Datenreport and Methodenreport papers are responsible for the contents of their respective articles.

---

## 1 Introduction

The IEBS is a random sample drawn from the **Integrated Employment Biographies (IEB)** of the IAB. The IEB are not to be understood as a self-contained dataset but as a procedure for merging data from four different sources for the purpose of data quality control and for drawing samples such as the IEBS. The four data sources are

- the **IAB Employee History (BeH)** with observations of employment subject to social security taken from the social security notification procedure,
- the **Benefit Recipient History (LeH)** with observations of receipt of unemployment benefit, unemployment assistance and maintenance allowance,
- the **Participants-in-Measures History File (MTH)** with observations of participation in employment and training measures and
- the **Applicant Pool Data (BewA)** with job-search observations.

The IEB were developed in the context of the Biographical Data and colAB projects and were supported by a committee of advisors from the IAB research departments. The IEBS 2008 is based on the IEB Version 7.02.

Important comments regarding data quality:

- IEB data always have the reservation that they are subject to later modifications by the institutions from which the data originate. Both during the social security notification procedure and in the BA working data corrections may sometimes be made even years after the end of an observation. The number of observations affected is small, however.
- IEB data are still at a comparatively early stage in their development. Experience shows that there are limitations in data quality for technical reasons. The data are released with this reservation.

Any inconsistencies occurring in the data can generally not be corrected clearly. The type of correction often depends on the research objective. In the IEB and the IEBS, known inconsistencies are therefore only corrected in unambiguous cases; in other cases they are merely documented. The amount of time and work involved in the data editing process is thus greater than it is for example in the IABS, which has been edited more extensively. Please report to the FDZ any irregularities or inconsistencies you find in either the data or this documentation.

Information on the IEBS and on data access can be found on the FDZ website (<http://fdz.iab.de>).

---

Changes and special features compared with the 2005 version of the IEBS:

The status of the data has changed considerably since the loading status of the IEBS 2005. In principle all notifications have been added immediately. In addition new values have been added to the employment status (in particular as a result of new cases of participation in measures) and the status after exit, as well as to the reason for exit. In some cases (see below) it was not possible to retain the previous values in full. The most important changes compared with the 2005 version of the IEBS are listed below:

1. Updating of the loading status: employment notifications up until 12/2007; notifications of participation in measures and job-search up until 09/2008 and information on benefit receipt until 10/2008.
2. Inclusion of new variables: marital status, applicant's skills level, number of children in the household, health impediments, classification of economic activities, willingness to travel/relocate, assignment to BA client group, type of termination of last job, number of employees in the employing firm, year of establishment of employing firm, desired working hours of the job sought, duration of remaining entitlement to unemployment benefit, planned start and end dates of participation in a measure, place of residence abroad
3. Discontinuation of previous variables: measure ID number<sup>1</sup>, correction indicator<sup>2</sup>, status of the end date<sup>3</sup>, continuation of unemployment after incapacity for work<sup>4</sup>, pension insurance institution<sup>5</sup>, contribution group<sup>6</sup>, start of unemployment<sup>7</sup>, duration of unemployment<sup>8</sup>, marginal part-time employment indicator<sup>9</sup>
4. The variable "grund" is recoded in the variable spectrum.
5. The missing values are recoded uniformly to the value -7 (unknown or missing information, error in the original value and allocation missing).
6. Reforms of district territories in Saxony-Anhalt and Thuringia result in new district numbers from 2007. This is to be taken into account when conducting evaluations at district level (MTH and BewA).

---

<sup>1</sup> The measure ID number is unsuitable as an identifier for a short-term incentive measure. Practical experience has shown that it is not possible to distinguish these from long-term projects.

<sup>2</sup> Following the IABS 2005 this variable is no longer recorded.

<sup>3</sup> Some of the end dates from the sources BewA and MTH can be assumed to be of a planning nature (this mainly concerns the measure groups FF and ESF). LeH end dates that are in the future (in relation to the loading status) can be assumed to be cases of right-censoring.

<sup>4</sup> This variable is no longer contained in the raw data for the IEBS.

<sup>5</sup> The information content of this variable was classified as insufficient/unsuitable for this dataset (cf. for example IABS and BAP, see <http://fdz.iab.de>).

<sup>6</sup> The information content of this variable was classified as insufficient/unsuitable for this dataset (cf. for example IABS and BAP, see <http://fdz.iab.de>).

<sup>7</sup> Can be generated by the user.

<sup>8</sup> Can be generated by the user.

<sup>9</sup> Can be generated by the user (covers employment codes 109 and 209 of the employment notifications).

---

### Special features

This dataset does not include any notifications in accordance with Social Code Book II.

As of 2004/2005 new administrative procedures (Verbis / CoSachNT) were introduced in the employment agencies. The introduction of the procedures was staggered across the regions. This may lead to changes in the data volume in the personal characteristics recorded by Verbis during this period and in the cases of participation in measures which are recorded via CoSachNT. As a result of the changeover some variables are no longer available or are only available in a different form. As far as possible this is pointed out again in the relevant places. When the CoSachNT procedure was introduced, the architecture of the data historisation for participation in measures was also changed. This resulted in an increase in the number of cases of participation, which had corresponding effects on the volume structure.

As of 2006 the administrative procedure for recording benefit receipt data, coLei, was switched over to the COLIBRI procedure (see 4.1.2 and 4.5.2).

## **2 Outline of the dataset**

The IEB are in principle equivalent to a census.<sup>10</sup> The IEBS sample is drawn by means of a random selection of eight birthdays. Every person born on one of these dates is selected, which is approximately equivalent to 2.2% ( $\approx 8/365$ ) of the individuals in the IEB. Without weighting, this birthday sample is representative with reference to the groups of persons (or data sources) included in the IEB at the relevant time. In other words: at any particular time the ratio of the stocks of individual person groups in the sample corresponds quite precisely to their ratio in the IEB and thus also to reality.<sup>11</sup> In this respect the IEBS differs from the IAB Employment Samples (IABS), where the sample is drawn by means of a random selection of employees' social security numbers.<sup>12</sup> Individuals who were not in employment subject to social security contributions in the period 1975 to 2001 are not included in the IABS – but they are in the IEBS.<sup>13</sup>

---

<sup>10</sup> Individual observations may be lost in the course of the data generation process. For example, observations that can not be related to a person are deleted (see Section 3.3). These lost observations are small in number, however.

<sup>11</sup> One restriction in the representativeness results from the fact that individuals whose date of birth was not known when they were assigned their social security number were given a certain date of birth which was not one of the eight designated birthdays for the sample. Most of these individuals are older foreigners, which means that this group is likely to be slightly underrepresented in the IEBS.

<sup>12</sup> The IABS sample is drawn using an east/west and a foreigner/German stratification.

<sup>13</sup> There are no frequency counts regarding this selectivity in the IABS data. However, the selectivity is likely to be of importance only in the first years, as previous employment subject to social security contributions is in principle a precondition for benefit receipt.

---

## 2.1 Overview

<b>Content characteristics</b>													
Topics	<p><i>Socio-demographic characteristics:</i> gender, year of birth, education, nationality, region type, severe disability status, school-based qualification(s), applicant's skills level, marital status, number of children in household, health impediments, place of residence and place of work</p> <p><i>Employment:</i> occupation, daily remuneration, occupational status and working hours, economic activity, transition zone</p> <p><i>Benefit receipt:</i> unemployment benefit, unemployment assistance, maintenance allowance</p> <p><i>Measures of active labour market policy:</i> type of measure: job-creation measures (ABM), general structural adjustment measures (SAM), promotion of vocational training and retraining (FBW), short-term training schemes (TM), German language courses (DSL), independent employment promotion measures (FF), joint programme of the Federal Employment Agency and the European Social Fund (ESF_BA), personnel services agency (PSA), programmes to get young unemployed people into employment and training (JUMP, JUMP_PLUS), commissioning of third parties to conduct job placement), periods of assistance, planned start and end dates</p> <p><i>Job search:</i> job-search status, employment status prior to job search, type of termination of last job, willingness to seek work anywhere in Germany, BA client group, desired working hours of the job sought, duration of remaining entitlement to unemployment benefit</p> <p><i>Establishment:</i> economic activity, no. of regular employees, year established</p>												
Data unit	Employees covered by social security (including those in marginal part-time employment since 1999), benefit recipients, jobseekers, participants in employment or training schemes												
Number of cases	1,487,835 individuals 23,667,828 original observations, 30,693,764 non-overlapping observations												
Period covered	The period covered depends on the data source. <sup>14</sup> <table border="1" style="width: 100%; border-collapse: collapse;"> <tbody> <tr> <td>Employment</td> <td>1/1990-12/2007</td> <td>(BeH)</td> </tr> <tr> <td>Benefit receipt</td> <td>1/1990-10/2008</td> <td>(LeH)</td> </tr> <tr> <td>Participation in measures</td> <td>1/2000-09/2008</td> <td>(MTH)</td> </tr> <tr> <td>Job search</td> <td>1/2000-09/2008</td> <td>(BewA)</td> </tr> </tbody> </table>	Employment	1/1990-12/2007	(BeH)	Benefit receipt	1/1990-10/2008	(LeH)	Participation in measures	1/2000-09/2008	(MTH)	Job search	1/2000-09/2008	(BewA)
Employment	1/1990-12/2007	(BeH)											
Benefit receipt	1/1990-10/2008	(LeH)											
Participation in measures	1/2000-09/2008	(MTH)											
Job search	1/2000-09/2008	(BewA)											
Time reference	Continuous employment history, recorded on a day-to-day basis												
Regional structure	Place of work: western/eastern Germany, federal states, regional directorates; Place of residence: western/eastern Germany, federal states, regional directorates (not for all data sources and periods) for other variables see under "sensitive variables"												
Territorial allocation	BeH: 31.12.2007 (updated district territories) LeH, MTH, BewA: original district territories; no corrections												
<b>Methodological characteristics</b>													
Survey design	approx. 2.2 % random selection drawn from the individuals included in the IEB												

<sup>14</sup> A more precise description of the periods covered can be found in Section 2.2.



Institutions involved in survey	Social security agencies, Federal Employment Agency
Frequency of data collection	Permanent, updates of data basis (IEB): every 9 months
File format and size	STATA (2.9 GB), SPSS and SAS on request
Data access	On-site use at the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB)

## 2.2 Volume structure and degree of completeness over time

No. of cases in the IEBS		before splitting	after splitting
BeH		16,817,514	19,152,561
LeH		2,869,572	4,420,896
MTH		569,130	1,029,709
of which	ABM, SAM etc.	145,830	331,056
	FBW, TM, DSL	298,302	457,309
	FF	42,749	77,450
	ESF-BA	12,584	26,138
	PSA	3,575	7,882
	JUMP	18,894	43,302
	JUMP-PLUS	2,917	5,785
	comm. of third parties	44,279	80,787
BewA		3,503,691	6,201,840
Total no. of observations		23,759,907	30,805,006
Individuals		1,487,835	

Note: for the distribution of the observations see the variable "Employment status: person group, type of benefit, type of measure, job-search status" in Section 4.5.2.

An episode splitting procedure is used when the IEBS is compiled. For this, where periods of time overlap within an account, the observations are cut and duplicated in such a way that only observations which are completely parallel and thus free of overlaps result. This leads to an increase in the number of observations. A detailed account of the episode splitting procedure is provided in Section 3.6.

The frequencies of the observations in the four data sources can be found in the frequency counts for the variables of the start and the end of the observation. There, inconsistencies in the degree of completeness over time are marked. These inconsistencies are explained in the following:

**BeH:** Due to the introduction of the employment notification procedure in the federal states of eastern Germany, the notifications for BeH observations can only be assumed to be sufficiently complete from 1993 onwards. The increase in the number of BeH observations from 1999 onwards is the result of the obligation to submit employment notifications for people in marginal part-time employment from 01.04.1999 onwards.

**LeH:** Start dates before 1990 can be explained by the fact that all observations which ended in 1990 or later are included. The few end dates before 1990 occur as a result of subsequent corrections. For the states of eastern Germany the observations are only complete from 1992 onwards. If the end date is 14.10.2008 or later it is generally the end of the period for which benefit was allowed, and not necessarily the end of benefit receipt.

**MTH:** Observations of participation in employment and training measures are complete from start dates in the year 2000 onwards. All observations which go into the statistics of the Federal Employment Agency from the year 2000 onwards and whose end date is 1995 or later are recorded. End dates before 2000 occur when staff at an employment agency enter or correct cases of participation in measures with a delay and as a result of errors made when entering the data. End dates after 2008 are planned end dates.

**BewA:** BewA observations are included from end dates in the year 2000 onwards. Older observations are only included for jobseekers who were recorded under the same client number both before the year 2000 and from 2000 onwards. Very early observations (before 1990) and end dates after 2008 are likely to be mainly typing errors, however.

---

## 2.3 Differences between the IEBS and the IABS and the BA Employment Panel

Various samples of the BA individual data are available which are suitable for different analysis purposes.<sup>15</sup> The following table compares the IEBS with the IABS and the BA Employment Panel datasets, which are partly based on the same data sources.

Characteristics	IEBS 2008	IABS-R75-04	BA Employment Panel
Period covered acc. to observation type:			
Employment	1990-2007	1975-2004	1998-2006
Benefit receipt	1990-2008	1975-2004	
Participation in measures	2000-2008		
Job search	2000-2008		
Survey design	2% sample	2% sample	2% sample
Time reference	non-overlapping observations	non-overlapping observations	reference date panel, quarterly, 36 waves
Subsequent corrections	are incorporated	are incorporated	are not incorporated (in some cases notifications are continued)
Correction of gaps and inconsistencies	corrections only in unambiguous cases	more extensive corrections to facilitate evaluation, e.g. generation of artificial employment observations	few corrections of inconsistencies, continued notifications are retained
Special features	observations of participation in employment and training measures and observations of job search	variables on the employment structure of the establishments, large user group, good knowledge of the data quality	variables on the employment structure of the establishments, aggregations correspond to the values of the BA statistics
Data access	on-site use	on-site use, SUF	on-site use, remote data access, SUF

## 3 Data preparation

### 3.1 Data sources

#### 3.1.1 Employee History of the IAB (BeH)

The Employee History of the IAB (Beschäftigten-Historik des IAB – BeH) contains the earnings notifications submitted by the establishments to the social security agencies in accordance with the Regulation on Data Collection and Transmission (Datenerfassungs- und -übermittlungsverordnung

<sup>15</sup> An up-to-date overview of the data available from the Research Data Centre of the Federal Employment Agency at the Institute for Employment Research can be found under <http://fdz.iab.de>.

- DEÜV).<sup>16</sup> The earnings notifications consist of annual, end-of-employment and employment interruption notifications and notification revisions for every employee covered by social security and, from 01.04.1999 onwards, also for employees in marginal part-time employment.

The period 1990 to 2007 of the BeH is used for the IEBS. From 1999 onwards the so-called annual employment data (Jahres-Zeitraum-Material - JZM) of the statistics goes into the BeH.

The annual employment data (JZM) of the statistics differ in the yearly data with regard to the periods taken into account for retrospective notifications:

- 1999-2003: 36-month version
- 2004-2005: 18-month version
- 2006: 12-month version (officially not released)
- 2007: 6-month file (officially not released)

A degree of completeness of just under 90% can be assumed for 2007. These data have already undergone various cleansing and correction processes. For more details on this see Section 3.3.

### 3.1.2 Benefit Recipient History of the IAB (LeH)

The period from 1990 up to the latest available data (14.10.2008) of the Benefit Recipient History (Leistungsempfänger-Historik des IAB - LeH) is used for the IEBS. The LeH contains the deregistrations submitted by the BA to the health insurance institutions following periods of benefit receipt, and the latest available data on notifications of current benefit receipt. In earlier versions the data originated solely from the old administrative procedure, coLei, and were supplemented with the latest available data from the Benefit Payment Control File (Zahlungsnachweisdatei - ZND). In the current LeH, from 2004 onwards the data come from the Data Warehouse (DWH), where the latest available data from the successor administrative procedure, COLIBRI, are recorded exact to the day.

The switchover from the coLei procedure to COLIBRI took place gradually from 2004 to 2006. From 01.07.2004 onwards the data were recorded using COLIBRI in the employment agency in Gotha. From December 2005 onwards the remaining employment agencies changed over to the new system. There was not a complete migration of the data, but instead the benefit records were removed one by one from the old system by the person responsible at the employment agency and were taken over into the new system. By 31.12.2006 the changeover was completed. The new benefits in accordance with Social Code Book II, which were introduced on 01.01.2005, are not

---

<sup>16</sup> The Regulation on Data Collection and Transmission (DEÜV) is documented extensively and clearly by various health insurance providers, see e.g. Ratgeber zur Sozialversicherung 2009, Broschüre des Deutsche BKK Arbeitgeberservice, [http://www1.deutschebkk.de/Service/Arbeitgeberservice/Wissen\\_von\\_A-Z/5B2D7F649D116FBDC1256DC90\\_0510B3E/RatgeberSV\\_2009\\_Internet.pdf](http://www1.deutschebkk.de/Service/Arbeitgeberservice/Wissen_von_A-Z/5B2D7F649D116FBDC1256DC90_0510B3E/RatgeberSV_2009_Internet.pdf), S.32 (link accessed on 04.05.2009).

recorded in COLIBRI. However, in 2004 it was still possible to record in COLIBRI payments of Unemployment Assistance and Maintenance Allowance, which were discontinued as of 01.01.2005. As the procedure was switched over at different times in different areas, it has an effect on evaluations of the LeH at regional level.

### 3.1.3 Participation-in-Measures History File (MTH)

The Participation-in-Measures History File (Maßnahme-Teilnehmer-Gesamtdatenbank – MTH) of the IAB contains data on employment and training measures in account form. Several files are produced from a complex database containing historicised information of all movements which are recorded statistically:

ABM	Measures of active labour market policy (job-creation measures (ABM)/ structural adjustment measures (SAM), but not one-off payments such as MOBI travel expenses)
FF	Independent employment promotion measures (Freie Förderung)
FbW	Further vocational training measures, short-term training measures (TM) and German language courses (DSL)
ESF-BA	Joint measures of the Federal Employment Agency and the European Social Fund (excl. the measure comprising social-education support)
PSA	Personnel service agencies
JG	Immediate action programme for the training, qualification and employment of young people
AfL/JP	Jobs for the long-term unemployed (AfL) / special programme to get young unemployed people into employment and training (Jump Plus)
P37	Commissioning of third parties to conduct job placement in accordance with § 37 of Social Code Book III (SGB III)
AGH	Work opportunities
ESG	Entry-wage top-up/business start-up allowance for recipients of Unemployment Benefit II in accordance with § 29 of the Social Code Book II (SGB II)
SWL	Other benefits/support in accordance with § 16 sub-section 2 sentence 1 of the SGB II

These files, in which every case of participation in an employment or training measure is linked with additional variables from the Applicant Pool Data (BewA) (e.g. severe disability status), constitute the MTH part of the IEBS.

The observations from the MTH which are used are complete for the period 1/2000 - 09/2008; some cases of participation in measures prior to 2000 are also included. Cases of participation in measures with a start date before 1995 were generally excluded. The degree of completeness of data concerning recent cases of entry into a measure can be assumed to be lower. Due to the change in the data source (IAB research database, ISAAK, instead of the MTG (participants-in-measures data), the degree of completeness for participation in measures in 2000 with a start date before 2000 can be assumed to be considerably higher than was the case in the IEBS 1.0.

As in the data source procedures several, or in some cases many, data records are supplied regarding a case of participation in a measure, and since there are no clear criteria in the data for identifying whether these data records concern single or multiple cases of participation, a control

system was developed and agreed upon for the IAB research database in order to identify cases of participation. Further information about this control system is available on request.

#### 3.1.4 Job-search status from the Applicant Pool Data (BewA)

The data regarding the job-search status are originally extracted from the DWH layer of the Applicant Pool Data (Bewerberangebot - BewA). In this database the job-search status and the job applicants are recorded under different identifiers; the job-search status under the client number, the job applicant under the BewA data entry number. In the IAB research database, all job applicants are assigned to their corresponding client numbers. In this database, too, the data are available as historicised data in account form. For the IEB the job-search data (Arbeitssuche - ASU) are drawn from the ASU evaluation file of the ISAAK application layer in which job-search observations are linked with client data and details from the Applicant Pool Data.

The data are available in full for the period from 1/2000 to 09/2008. If jobseekers were recorded under the same client number prior to the year 2000, then these data are also included.

### 3.2 Linking the data

The four data sources are linked via the social security numbers and the BA client numbers. Various problems arise here. A multitude of employees never come into contact with the BA and accordingly have no BA client number. Some employees have more than one social security number at times. This is corrected by the Association of German Pension Funds (Verband deutscher Rentenversicherungsträger - VDR) as soon as the situation is discovered. The VDR compiles a correction file, which is incorporated into the BA and IAB data by applying one social security number to all of a person's observations. For benefit recipients, both the social security number and a client number are generally available, whereby a person frequently has more than one client number. It is only since 1999 that unique client numbers have been allocated, and since 2003 the allocation of client numbers has been coordinated across Germany using the Central Data Management Procedure for Data on Individuals (zentrale Personendatenverwaltung - zPDV). The MTH and the BewA often only contain the client numbers.

The following linkage rules apply for the IEBS:

From the client information in the Data Warehouse (DWH) of the BA the client number – social security number relation (KNR-VSNR) is established, which contains precisely one social security number (VSNR) for each client number (KNR). The following rules are observed when establishing this relation:

- for each client number the "most up-to-date" social security number is found,

- the date of birth and gender from the client information in the DWH are added to each client number, and the date of birth and gender from the social security information in the DWH are added to each social security number,
- the social security number is then rejected if the first name **and** the surname **and** the date of birth differ in the social security data and the client data,
- then the VDR file containing the cancellations of social security numbers is incorporated,
- if the social security number was changed as a result of incorporating this file, then the date of birth and gender have to be re-calculated.

For records from the sources LeH, LHG, MTH and ASU the social security number, the date of birth and the gender from this relation (KNR-VSNR) are merged if the client number is available.

LeH data are originally organised according to BA client numbers. The social security number is also contained in a very large percentage of the data (99.7 %), however. For LeH observations, the relation comprising client number and social security number, if available, is used for observations with an end date of 01.01.1999 or later. For observations before 1999 it is only used when the combination of client number and social security number occurring there also appears in a dataset from 1999 onwards and when a social security number is found for this client number in the relation. If it is not possible to establish unambiguity in the LeH observations from 1999 onwards via the relation, i.e. if the client number is not contained in the relation and if there is more than one social security number for the client number in the LeH, the social security number is deleted in the accounts.

In the case of MTH and ASU data, a person is originally identified in ISAAK via the BA client number (approx. 80 %) or the BewA data entry number (approx. 20 %). However, a client can have several different client numbers e.g. after moving house during the time before the Central Data Management Procedure for Data on Individuals (zPDV) was introduced in 2003. For a large percentage of the data (MTH: over 90 %; ASU: over 80 %) it is possible to assign a social security number via the client number or the BA client data; for a very small percentage of records (MTH: below 1 %, ASU: approx. 7 %) neither a social security nor a client number is available.

The quality of the assignment of social security numbers to client numbers was tested when the relation was introduced:

- Of approx. 900,000 clients those were selected who had the same social security number but different first names or surnames. 557 such data records were found.
  - Of 29,465,474 cases relating social security numbers to client numbers, approx. 8,200 cases were detected in which the individuals concerned were presumably different people. These cases are excluded when the relation is created by means of the condition "first
-

name or surname or date of birth must be identical in the social security data and the client data".

On the whole it can be established that the relation of client-person covered by social security in the DWH aggregates rather too much, but is of quite good quality.

### **3.3 Corrections and filters**

#### **3.3.1 BeH**

The BeH observations have undergone the following correction steps:

- Minor corrections were made to dates.
  - In the annual employment data (JZM) from 1999 onwards only the person groups 101, 102, 103, 105, 106, 112, 118, 119, 120, 140, 141, 142, 143, 201, 203 and 205 (from the BA statistics) are supplied, as well as the two groups 109 and 209, which indicate people in marginal part-time employment. Groups that are not supplied include, from 1999 onwards, people in short-term employment, i.e. person groups 110, 202 and 210.
  - The BeH prior to 1999 contains all person groups, including those that are no longer supplied with the annual employment data from 1999 onwards.
  - When distinguishing between employees subject to social security and people in marginal part-time employment, person groups 109, 110, 202, 209 and 210 are classified as "in marginal part-time employment".
  - Registrations are filtered out.
  - Cancellation notifications and revision notifications are incorporated, cancellation records are deleted.
  - Information on the establishment number (e.g. place of work, economic activity) is added.
  - Duplicates are removed according to the following criteria: social security number, establishment number, start and end dates, reason for notification, remuneration and date created.
  - Data records from the BeH are only taken into the IEB for the period 1990 to 2007.
  - Social security numbers are updated to the status as of 31.12.2006; cancelled social security numbers are deleted.
  - Gender and date of birth are taken from the social security number.
  - The territorial allocations for place of work and place of residence are updated to the status as of 31.12.2007.
-



### 3.3.2 LeH

For LeH observations the following correction steps and filter conditions are carried out. It is possible to distinguish between basic corrections and other corrections.

The following procedures come under the basic corrections:

- Filtering out of contributions to health insurance / long-term care insurance (§207 A SGB III)
- Date corrections
- Merging of the social security number via the client number from the client number – social security number relation (including correction of the cancelled numbers as provided by the Association of German Pension Funds (VDR)). LeH records prior to 1999 that do not have a social security number are excluded, as in these cases the client number can not be assumed with certainty to be correct.

These basic corrections as well as other corrections are listed in detail below:

- For the IEB, only observations from the LeH for the period from 1.1.1990 (end date) until the latest available data as of 11.09.2008 are taken.
  - It should be taken into account that in cases of ongoing receipt the end date corresponds to the end of the period for which the benefit is allowed and not the end of the benefit receipt period. Observations with no end date or an invalid end date are excluded from further processing, since according to information provided by the specialists, in the case of records without an end date it can not be assumed that a benefit payment was made at all.
  - Observations without a valid start date are excluded from further processing.
  - Observations whose end date precedes the start date are excluded.
  - If the end date for unemployment assistance precedes the start date by one day, then the end date is deferred by one year.
  - Only observations with the grouped benefit types 1 (unemployment benefit – ALG), 2 (unemployment assistance - ALHI), 3 (maintenance allowance - UHG) are taken into the IEB. The values 5 (§166b Employment Promotion Act (AFG) / §207a SGB III) and 9 (miscoding) are filtered out.
  - Only formally correct social security numbers are included (according to checking code "not deleted" and "correct").
  - Social security numbers are updated to the status as of 30.06.2006; cancelled numbers are not included.
  - The territorial allocations are not updated, as only the employment agency of the place of residence is available and not the local authority, which would be required for this.
-

- Correction of overlapping LeH observations

Overlapping LeH observations are corrected for the IEB in accordance with the following rules:

- All overlaps in which both of the observations involve **the same type of benefit** and the same amount of benefit are corrected as follows:
  - If one observation is completely embedded within another one, the shorter observation is deleted. If the two observations are of the same length, one (randomly selected) record is deleted.
  - In cases where observations partially overlap, the end date of the observation that began first is corrected.
- Overlaps of observations involving **different types of benefit** and/or different amounts of benefit paid are only corrected if the overlap concerns just one day and the original duration of both observations exceeds one day. In these cases the end date of the observation that began first is put back by one day.

### 3.3.3 MTH

For the IEBS the MTH data from the ISAAK application layer are used, which consist of evaluation files for the sub-categories ABM/SAM, FbW, FF, ESF-BA, PSA, JG, AFL//JP, P37, AGH, ESG and SWL. Here the following correction steps and filter conditions are carried out:

- The data from the MTH which are used cover the period from 01/1995 to 09/2007. Data records with a start date before 1995 were deleted. The cases of participation in measures are only available in full from 01/2000 (start date of the measure) onwards, however.
  - If the client number is available, the social security number is taken from the KNR-VSNR relation. If the client number is not available, the social security number is taken from the BA client data (HIST-ISAAK-Person).
  - Cancelled or changed social security numbers are updated from the VDR file to the status as of 30.06.2006. Records with cancelled social security numbers are retained if a client number is available; the social security number is set to 'missing'.
  - Gender and data of birth are taken from the social security number if this is possible.
  - The territorial allocations are not corrected.
  - The region type 06 is generated via the current relation of employment agency – region type (AA-Regiotyp) of the DWH.
  - For the provision of the ESF data from the MTH for the IEB, all social-education measures are filtered out as they are not measures but benefits and the data contain only a start date and no duration.
-

- In the case of the data from the ABM category, all one-off benefits (MOBI travel expenses, and careers advice and placement allowances (Unterstützung der Beratung und Vermittlung - UBV)) are excluded.
- Data records are filtered out if neither social security number nor client number are available, as it is not possible in such cases to identify the record or to allocate it to an account.
- Records whose end date precedes the start date are filtered out.
- Accounts with more than 99 observations are not included in the IEB.

### 3.3.4 BewA

For the IEB the following correction steps and filter settings are carried out on the data from the job-search (ASU) evaluation file of the ISAAK application layer:

- The job-search data (ASU) that are used cover the period from 01/1997 to 09/2007. From 1999 onwards the full data are contained.
  - If the client number is available, the social security number is taken from the KNR-VSNR relation. If the client number is not available, the social security number is taken from the BA client data (HIST-ISAAK-Person).
  - Cancelled or changed social security numbers are updated from the VDR file to the status as of 30.06.2006. Records with cancelled social security numbers are retained if the client number is available. In these cases the social security number is set to 'missing'.
  - Gender and data of birth are taken from the social security number if this is possible.
  - The territorial allocations are not corrected.
  - Generation of employment status: the data basis only distinguishes between records with the status "unemployed" and "jobseeker". In the IEB a distinction is made between "unemployed", "incapacitated for work" and "jobseeker". Records with the employment status "incapacitated for work" are data records for which (1) a previous unemployment record exists which joins the next record without a gap and has "incapacitated for work" as the reason for exit (sna [status after exit] = 6004), (2) a follow-up unemployment record exists which also follows without a gap and (3) the record itself does not have the status "unemployed" but "jobseeker".
  - The region type 06 is generated via the current relation of employment agency – region type (AA-Regiotyp) of the DWH.
  - Data records are filtered out if neither social security number nor client number are available as it is not possible in such cases to identify the record or to create an account.
  - Records whose end date precedes the start date are filtered out.
  - The accounts are created according to social security number and client number (social security number takes precedence).
-

### 3.3.5 Aggregate file

The IEB aggregate file is created from the data sources for employees (BeH), benefit recipients under Social Code Book III (LeH), benefit recipients under Social Code Book II (LHG), participants in measures (MTH) and jobseekers (ASU/BewA). The following data processing steps are conducted:

- Creation of accounts via social security number or client number, whereby the social security number always takes precedence.
- Allocation of a technical account ID as an identifier. From IEB V5.00 onwards the account IDs allocated in the previous version with their relation to the social security and client numbers are retained as account-forming codes.
- Data records with neither a client number nor a social security number are deleted.
- Data records in which the age is under 13 or over 75 are deleted.
- Records whose end date precedes the start date are deleted.
- Accounts in which the gender or date of birth change are deleted.

### 3.4 Legally inadmissible parallel states

In the IEB, data are combined from different operational systems which are not compared with one another. In order to correct legally inadmissible parallel states it is necessary to have extensive knowledge of the relevant legislation. A particularly useful overview of admissible and inadmissible states is provided in Fitzenberger et al. 2005b. The results of a project to monitor inconsistent states are described in Jaenichen et al. 2005. Here we provide a selection of introductory comments:

- Inconsistencies may occur as a result of errors made during the data entry process. If client numbers have been entered incorrectly, accounts might not be merged correctly and thus parallel records which are actually expected may be missing.
  - In addition it can happen for cases of participation in a measure that an updated observation is not recorded if the person concerned drops out of the measure, with the result that a spell of unemployment benefit receipt already appears in the data again parallel to the participation in a measure.
  - The data in the different operational systems might not be recorded on the same day but a few days apart, so for example a notification of exit from unemployment is only made several days after the beginning of an employment or training measure.
  - From 1999 onwards there are job-search (ASU) records parallel to unemployment benefit records from the LeH.
-

- There are not necessarily any LeH records parallel to Applicant Pool Data (BewA) records if the jobseeker is not entitled to benefits.
- There may also be BeH records parallel to BewA records with unemployment status or LeH records.
- In the case of pure FbW measures the participants should not be registered as seeking work at the same time. There need not therefore be any parallel BewA records for the entire period. The participants do not register as seeking work again until just before the end of the measure. Parallel BewA records are then possible until the end of the measure.
- Further vocational training (FbW), short-term training courses (TM) and German language courses (DSL) generally have parallel observations of maintenance allowance (UHG) in the LeH. Due to an amendment in the legislation, from 01.01.2004 onwards participants in short-term training courses are no longer counted as unemployed. It has not yet been investigated whether this has an effect on the data. Exception: the participant is taking part in an ESF-BA measure parallel to a FbW measure. In this case there must not be an observation of maintenance allowance in the LeH.
- In the case of TM and DSL measures the participants are registered as jobseekers and may also be registered as unemployed at the same time.
- Participants in job-creation measures (ABM) and structural adjustment measures (SAM) should be registered as jobseekers but not as unemployed.
- In the case of ABM and SAM measures there should also be parallel employment records.

### 3.5 Data quality and problems

The account of the problems involved in linking the data, the data corrections made to date, and the reference to the occurrence of legally inadmissible parallel states given in the previous sections provide a sense of the quality of these process-generated data. Corrections were only made by the IAB if they were considered to be of general value. Other known quality problems are:

- Inconsistencies occur as a result of errors made during the data entry process. If client numbers have been entered incorrectly, accounts are not merged correctly and expected parallel records are therefore missing.
- It sometimes happens that an updated observation is not recorded if a person drops out of a measure, with the result that an observation of benefit receipt already appears in the data again parallel to the participation in a measure.

The users of the data are called upon to add to the list of inconsistencies should they find any and to suggest corrections or data cleansing procedures.

In contrast to the statistical data published by the BA and the BA Employment Panel, in the IEBS observations which are recorded or corrected at a later date are taken into account, which can not

---

be done in statistical data that is published in a more timely manner. For this reason it is not possible to reproduce the figures from the statistical data provided by the BA using the IEBS. Especially when new measures are introduced, considerable discrepancies to the figures published by the statistics department of the BA are to be expected due to observations being recorded retrospectively.

### 3.5.1 Employee History (BeH)

#### Employees in marginal part-time employment

The BeH data records have a peculiarity regarding individuals in marginal part-time employment. The obligation to submit employment notifications for people in marginal part-time employment (or to be more precise: the transmission of these data to the BA) has been in existence since 01.04.1999. Prior to this date the BeH contains no data about people in marginal part-time employment. In the data it then looks as if these marginal part-time jobs began on 01.04.1999 (ieb\_beg\_orig = 1.4.1999). Presumably a not inconsiderable number of the cases which have 01.04.1999 as the start date are jobs that had been in existence for a longer time (also since the previous year) and should actually have 01.01.1999 as the start date. The obligation to submit employment notifications for marginal part-time employees results in a substantial increase in the total number of BeH data records in 1999. Furthermore there are more overlaps from 01.04.1999 onwards – both within BeH records and with other data sources – since for example marginal part-time employment and benefit receipt are possible simultaneously.

#### Degree of completeness of the education variable

Analyses reveal that the number of missing values for the education variable increases over time. This becomes clear in frequency counts of first notifications by years. In the latest available data the percentage of missing values for the BeH data source is over 30%. Further evaluations give rise to the suspicion that establishment size effects may also play a role here (larger proportion of missing values in small establishments).

### 3.5.2 Benefit Recipient History (LeH)

#### Introduction of the Colibri administrative procedure

In 2004 (pilot scheme in the employment agency in Gotha) and 2005 a new administrative procedure to record periods of benefit receipt was gradually introduced. A particular problem which can be found in the data concerns the transitional phase, i.e. the periods during which data records are

---

available in the LeH and in the DWH (01.01.2003 until 31.12.2005). Owing to the switchover from the old procedure to the new one and the overlapping recording periods in the old LeH and the DWH, it is not possible to incorporate the data into the new LeH according to fixed validity periods. This can lead to the number of benefit receipt notifications being underestimated during the transitional period.

### 3.5.3 Participants-in-Measures History File (MTH)

#### Alteration in the data generation (MTG / MTH)

In 2004 the coSach administrative procedure was superseded by a new procedure which is characterised in particular by an improved linkage to the DWH of the BA. The rules which were used to set up the MTG (participation-in-measures data) as a data basis for participation in measures can no longer be retained after this change in the administrative procedure. The MTG is superseded by the MTH (participation-in-measures history file). The rules and checking procedures used in the MTH provide for an increase in the figures for participation in measures of active labour market policy (see also Engelhardt et al., 2008).

#### Start and end dates

In MTH data records the start and end dates were originally recorded by administrative assistants and were not checked for plausibility by a program in the old coSach procedure. It can be assumed that in a small percentage of the data these dates are incorrect. As implausible details appear in individual data records, e.g. 01.01.1990 as the start and the end date, or a start date in 1920, all data records with a start date before 01.01.1995 were excluded in general.

In the coSach NT procedure, which has been in use since 21.12.2004 for all coSach sub-categories, plausibility checks are already conducted when the data is entered wherever is possible.

#### Known problem in the case of the bridging allowance

In participation-in-measures records concerning the bridging allowance (Überbrückungsgeld - ÜG) there is a not inconsiderable number of cases with a duration of two years although this allowance may only be granted for a maximum of six months. The reason for this is that in coSach ABM the data entry mask has 01.01 of the respective year as the default setting for the start date and 31.12 of the following year as the default setting for the end date.

### 3.5.4 Applicant Pool Data (ASU / BewA)

#### Introduction of the Verbis administrative procedure

From mid-2005 to mid-2006 the coArb procedure, from which the applicant pool data originate, was superseded by the VerBIS procedure. In July 2005 the coArb operative system was first replaced by VerBIS in the employment agency in Wiesbaden as a pilot project. From December 2005 onwards it was then gradually replaced in several stages in all employment agencies. For this changeover procedure, the previous data from coArb were booted and migrated to VerBIS for each employment agency at the time of the changeover. The data from VerBIS are, however, "migrated back" into the data model developed for CoArb for the DWH (for the time being). Mapping the VerBIS data onto the previous variables in the DWH in this way leads to incomplete ASU observations which are characterised by a large number of missing values. At times there can even be a complete loss of ASU observations.

#### Incompleteness of the ASU observations before 2000

The data are available in full for the period 01/2000 to 09/2007 provided that jobseekers were recorded under the same client number before the year 2000. The completeness of accounts which go back further than the year 2000 is problematic, however.

The IEB data source ASU was set up with stock data from 1997 onwards. For this period different identifiers are used, which means that gaps may exist in the ASU notifications per person due to allocation problems:

- Accounts containing solely ASU observations before the year 2000 are incomplete with regard to the ASU observations as they may not have been allocated a client number (approx. 1% of all accounts in the IEB with ASU observations).
- Accounts which include observations with ASU periods before 2000 and also contain additional observations that extend into the period 01/2000 to 08/2002 or are entirely within this period are complete with regard to the ASU observations, however (approx. 1/3 of all accounts in the IEB with ASU observations).
- Accounts containing solely ASU observations that begin after 01/2000 are also complete with regard to the ASU observations (approx. 2/3 of all accounts in the IEB with ASU observations).
- Accounts that contain ASU observations before 2000 and after 08/2002, but none in the period 01/2000 to 08/2002 are not complete for the period before 2000 as there is no possibility to link observations before 2000 with later observations (less than 1 % of all accounts with ASU observations).
- For accounts that contain only ASU observations after 8/2002 it is not known whether there were also ASU observations before 2000 (approx. 30 % of all accounts with ASU observations).

#### Differing increases in the number of jobseekers in different regions

---



The introduction of the sphere of Social Code Book II (SGB II) into the active labour market policy of the BA was associated with changes in the organisation of responsibility (employment agencies working in cooperation with local authorities (ARGE), employment agencies and local authorities with separate SGBII tasks ("getrennt Trägerschaft") and local authorities being authorised to implement SGBII independently ("optimierende Kommunen")) and different data recording systems. In the administrative systems of the BA and in the research datasets of the IAB both a change in the institution responsible for a client and the allocation of jobseekers to institutions responsible for the implementation of SGB II are only depicted to a limited extent in longitudinal sections and are also only eligible for statistical analysis to a limited extent. In some cases this can be seen in regional differences in the number of jobseekers between the end of 2004 and the beginning of 2005. Here is an example: individuals who at the end of 2004 show job-search notifications for the sphere of SGB III with a validity until mid-2005 may be recorded in the sphere of SGB II from 1 February 2005 onwards. However, this is not recognisable in the data. The job-search notification under SGB III would then actually only be valid until the end of January 2005.

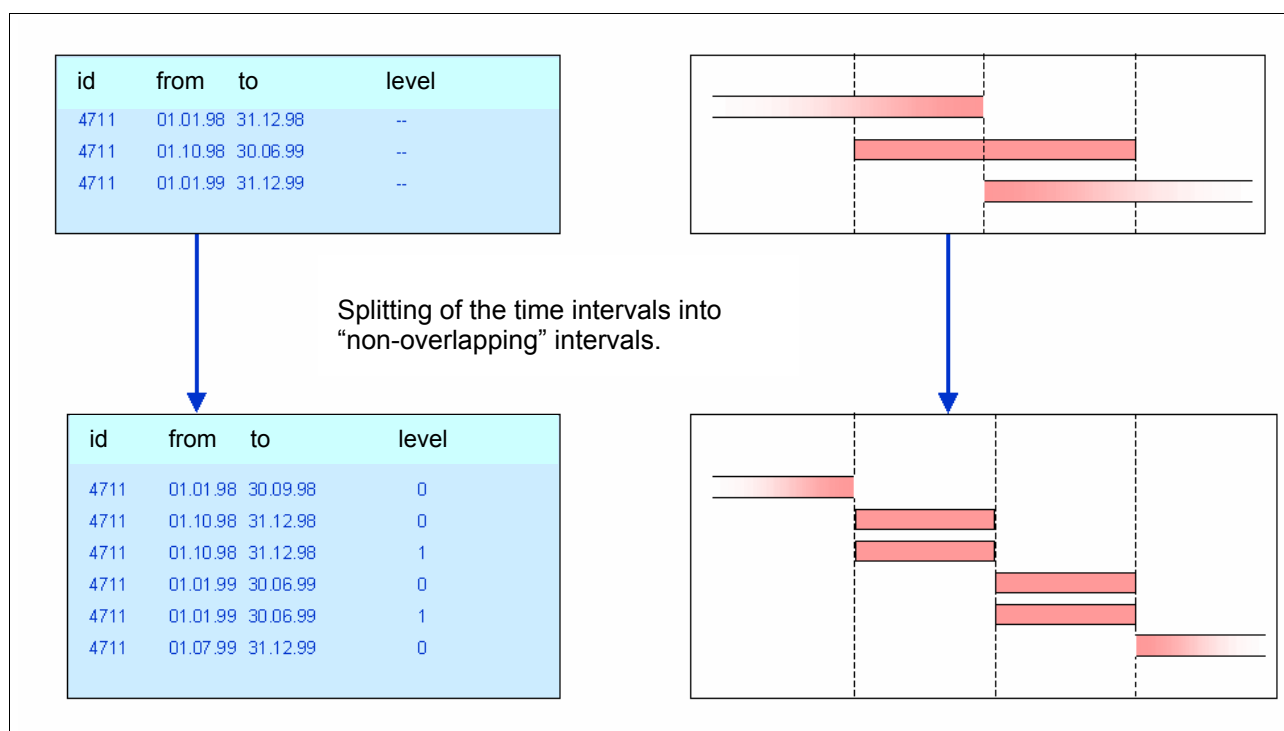
### 3.6 Episode splitting

When the aggregate file is compiled, episode splitting is performed. For this, where periods of time overlap within an account these observations are cut and duplicated in such a way that completely parallel periods and non-overlapping periods are created. This increases the number of observations.

Episode splitting results in observations being doubled and the period of validity being changed. 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`). Split episodes which were originally combined in one observation also have the same data record number.

There are various ways to restore the original data without the split episodes or to delete the episodes that were created artificially by means of episode splitting. These are by selecting

- all observations for which: start of the original observation = start of the split episode,
- all observations for which: end of the original observation = end of the split episode, or
- any observation per data record number.



For variables which are constant within the data record number due to their relationship to the original observation (see the line 'time reference' in the descriptions of the individual variables), it makes no difference which episode is selected per data record number. In the case of variables which are generated after the episode splitting procedure, which vary within the data record number, it does indeed make a difference. If one of these variables is to be used after the split episodes have been deleted, the criterion 'beginning of the original observation = beginning of the split episode' has to be used.

If the observations created by means of episode splitting are not deleted, the following must be taken into account: three variables refer to the state before the beginning of the original observation and therefore only apply to the beginning of the first of several split episodes:

- Employment status prior to job-search (estatvor)
- type of termination of last job (art\_kuend)
- planned start date (begplan)

Four variables refer to the state after the end of the original observation and therefore only apply to the end of the last of several split episodes:

- Reason for end of observation (grund),
- Status after exit (sna),
- duration of remaining entitlement to unemployment benefit (restanspruch)
- planned end date (endplan)

### 3.7 Anonymisation of data and sensitive variables

According to §75 of the Social Code Book X (SGB X) it is possible to transmit social data including personal identifiers if this is necessary in order to fulfil the research objective. The principle of data parsimony demands that no more information may be made available than is required to achieve the research purpose. It is always necessary to prevent the abuse of data (de-anonymisation, analyses that were not planned by the legislator etc.). In this weakly anonymised (pseudo-anonymised) IEBS this is done firstly by restricting access to on-site use (and remote data access in the medium term) and the organisational and technical measures associated with this.

Furthermore, the data are subjected to the following minor modifications: identifiers such as social security numbers, client numbers and establishment numbers are replaced by artificial numbers. The year of birth is included instead of the date of birth.

Certain variables which make it easy to de-anonymise individuals are only passed on in their original form if this is necessary for the analysis objective and if the reasons are explained in detail in the request for data access<sup>17</sup>. In the IEBS the variables which are particularly sensitive from the viewpoint of data protection legislation are:

- nationality
- the place of work and place of residence at district level
- the place of work and place of residence at employment agency level
- place of residence abroad
- state of health

### 3.8 Test data and comments regarding data evaluations

Test data are made available so that data users can prepare for their research visit and use their time at the FDZ efficiently. The test data are random variables which depict the univariate distributions of the variables and selected correlations in the data. They can be used to prepare and test evaluation programs but definitely not make content-based evaluations.

---

<sup>17</sup> Up-to-date details regarding data access can be found under <http://fdz.iab.de>.

---

## 4 Description of variables


The description of variables is divided into a general part comprising an overview of the variables and a description of the typification of the missing values that occur in the data. This is followed by descriptions of the individual variables, sorted according to contents. In this way similar variables are described in sections which are close together, sometimes even in the same section, with the aim of avoiding redundancy and making relationships between the variables clearer.

### 4.1 General part

#### 4.1.1 Overview of variables

The overview of variables lists the variable names and the longer descriptions of the variables. It also shows which variables occur or are available for the observations of the four data sources:

 available in the observations of this data source,

 not available in the observations of this data source, the variable contains solely system-missing values here.

Example: the planned start date of a measure is only available for MTH records, the observations of the other data sources contain system-missing values for this variable. Another characteristic of the IEBS is that some variables have different contents depending on the data source. Thus, for instance, for BeH observations the employment status variable contains the person group of the social security notification procedure, for LeH observations it contains the type of benefit, for MTH observations it is the type of measure and for BewA observations the job-search status. These differences are not immediately obvious from the variable name for every variable. A list of variables with the variable names in alphabetical order can be found in Appendix 6.2.

Variable name	Page	Description	BeH	LEH	MTH	BewA
<b>Identification numbers</b>						
persnr	31	Individual ID				
satznr	32	Data record number				
betnr	32	Establishment number				
<b>Start and end dates of observation</b>						
begorig	32	Original start date of observation				
endorig	32	Original end date of observation				
begepi	33	Start date of split episode				
endepe	33	Start and end dates				
<b>Generated technical variables</b>						
quelle	34	Source of the information				
kom_quel	34	Source combination				
spell	35	Observation counter per person				
nspell	35	Number of observations per person				
level2	36	Observation counter per episode				
nlevel2	36	Number of simultaneous observations per episode				
level1	36	Observation counter per episode and source				
nlevel1	37	Number of simultaneous observations per episode				
<b>Personal status before, during and after the current observation</b>						
estatvor	38	Employment status prior to job-search				
erwstat	38	Employment status: person group, type of benefit, type of measure, job-search status	from 1999			
grund	39	Reason for end of observation				
sna	40	Status after exit				
<b>Personal variables</b>						
gebjahr	41	Year of birth				
sex	42	Gender				
nation	42	Nationality				
schweb	42	Severe disability status				
schbild	43	School-leaving qualification				
		Notes	This information is required in accordance with the Schwerbehindertengesetz). A disability is defined as a permanent and severe from a 50% degree of disability. This variable is affected by the Verbis procedure (see Section 4.5.4). The variable is available in the data quality.			
bild	43	School education and vocational training				
quali	44	Skills level				
famstand	45	Marital status				
kind	45	Number of children in the household				
gesund_ein	46	Health				

Table continued on next page

Variable name	Page	Description	BeH	LEH	MTH	BewA
<b>Data on the employment relationship and job-search</b>						
stib	47	Occupational status and working				
beruf	50	Occupation				
tentgelt	48	Daily wage / daily benefit rate				
gleitz	49	Transition zone	from 03			
w73	49	Economic activity 73				
w93	49	<b>Fehler! Verweisquelle konnte nicht gefunden werden.</b> Economic activity 93	from 99			
w03	49	Economic activity 03				
mobil	50	Willingness to seek employment throughout Germany				
kunden_gr	51	BA client group				
art_kuend	51	Type of termination of last job				
arbzeit	52	Desired working hours of the job sought				
restanspruch	52	Duration of remaining entitlement to unemployment benefit				
begplan	52	Planned start date				
endplan	53	Planned end date				
<b>Data on the employing establishment</b>						
az_hpt	53	Number of regular employees				
grd_jahr	54	Year when establishment was founded				
<b>Location data</b>						
ao_bula	55	Place of work: federal state (Bundesland), district (Kreis)				
wo_bula	55	Place of residence: federal state (Bundesland), district (Kreis)	from 99			
ao_rd	56	Place of work: regional directorate (employment agency area)				
wo_rd	57	Place of residence: regional directorate (employment agency area)	from 99			
wo_kreis	55	Place of residence: district (Kreis)				
ao_kreis	55	Place of work: district (Kreis)				
ao_aa	56	Place of work: employment agency area				
wo_aa	58	Place of residence: employment agency area				
wo_aatyp06	58	Place of residence: employment agency region type 06	from 99			
wo_ausl	58	Place of residence abroad				

### 4.1.2 Missing values

The BA data do not have a standardised coding of missing values. The following standardisation was conducted in the IEBS:

Term	Value	Description	Example
No details available and no allocation possible	-7	In the original codes of the data sources, a value is often already stipulated for missing information following data collection. In this sense this category is also applicable when cells in observations are empty.  Values marked with -9 in the previous version are now also coded as -7. These include values which do not fit into the given categories and were also presumably caused by data entry errors or missing information.	Workers without further information about their occupation in the classification of occupations.
System-missing value		If a variable is not available for a data source (grey cells in the overview of variables in the previous section).  In dates, as negative values would be shown as dates.	

## 4.2 Identification numbers

### 4.2.1 Individual ID

Variable name	persnr
Variable label	individual ID
Available for	all sources
Origin	generated
Time reference	individual-related
Anonymisation	none
Detailed description	The individual ID is an artificial identifier for a person. The original identifier is mainly formed from the social security number. If no social security number is available, then the BA client number is used instead. Records which have neither a valid social security number nor a valid BA client number are rejected as they can not be linked. This identifier is replaced by a unique generated random number for every person.
Notes	The social security and BA client numbers are not transmitted for data protection reasons. This dataset thus meets the requirements of weak anonymisation.  See Section 3.1 on the uniqueness of social security and BA client numbers and the resulting problems when linking observations from different data sources.

#### 4.2.2 Data record number

Variable name	satznr
Variable label	data record number
Available for	all sources
Origin	generated
Time reference	variable within the individual ID
Anonymisation	none
Detailed description	The data record number is distinct across all of the original records of the IEB sources – also across the non-split records of the aggregate IEB. The variable 'satznr' allocates a unique number to every original observation. This makes it possible to merge further variables from other IAB data sources. New observations created by means of episode splitting retain the same data record number as the original observation.
Notes	none

#### 4.2.3 Establishment number

Variable name	betnr
Variable label	establishment number
Available for	BeH
Origin	BeH
Anonymisation	artificial identifier, only valid for the IEBS
Time reference	variable within the individual ID, constant within the data record number
Detailed description	Establishments with employees covered by social security are allocated establishment numbers by the employment agencies. In the IEBS the original numbers were replaced by artificial establishment identifiers. Every establishment is recorded under a unique identifier.
Notes	As the establishment characteristics of place of work and economic activity are merged via the establishment number, this information is not available in observations without an establishment number.

### 4.3 Start and end dates of observation

#### 4.3.1 Original start date of observation

See Section 4.3.4

#### 4.3.2 Original end date of observation

See Section 4.3.4



### 4.3.3 Start date of split episode

See Section 4.3.4

### 4.3.4 Start and end dates

Variable names	Original observations: begorig, endorig Split episodes: begepi, endepe
Variable labels	Original start date of observation (begepi: start date of split episode) Original end date of observation (endepe: end date of split episode)
Available for	all sources
Origin	BeH, LeH, MTH, BewA
Anonymisation	none
Time reference	begorig, endorig: variable within the individual ID, constant within the data record number begepi, endepe: variable within the individual ID and the data record number
Detailed description	begorig and endorig indicate the start and the end of the original observation, exact to the day, begepi and endepe indicate the start and the end of the non-overlapping split episodes.  In the case of MTH observations the end date of the original observation is taken from the exit data record provided that such a record is available. Otherwise it is taken from the entry data record.  Like all dates in the IEBS, the start and end dates are stored as integers. Here the date 31.12.1959 assumes the value -1, 01.01.1960 the value 0, 02.01.1960 the value 1 etc.
Notes	If due to overlaps an observation was divided into several episodes by means of episode splitting, then the beginning of the original observation corresponds to the start of the <u>first</u> split episode and the end of the original observation corresponds to the end of <u>last</u> split episode. All other start and end dates of the original observations and the split episodes differ from one another. If an original observation has not been split, the start and end dates of the episode correspond to the start and end dates of the original observation. If analyses are to refer to the unsplit observations, a filter can be set, for example using the condition begorig = begepi. Split episodes which were originally combined into one observation have the same data record number (satznr). A detailed description of the episode splitting procedure can be found in Section 3.6.  Attention: Some variables are only valid for the original observations and not for the observations added by means of episode splitting (see Section 3.6).  The episode splitting procedure results in episodes which are already completed before the observation period of a data source or do not begin until after the observation period of a data source.  In cases of overlaps it must be taken into account with regard to information on dates that the validity of the dates varies with the data source. In the BewA source, for example, observations before 1990 appear. These dates are presumably typing errors. Such early (or late) observations, also outside of the actual period of validity in the other data sources, are not corrected in the data.

## 4.4 Generated technical variables

### 4.4.1 Source of the information

Variable name	quelle
Variable label	source of the information
Available for	all sources
Origin	generated
Time reference	variable within the individual ID, constant within the data record number
Anonymisation	none
Detailed description	Contains the details as to the data source from which the information about the observation is taken. In the case of participation-in-measures data (MTH) a differentiation is made according to the origin from the specific participation-in-measures history, which sometimes, but not always, corresponds to the differentiation in the sub-categories of coSach.
Notes	none

### 4.4.2 Source combination

Variable name	kom_quel																									
Variable label	source combination																									
Available for	all sources																									
Origin	generated																									
Anonymisation	none																									
Time reference	variable within the individual ID and the data record number																									
Detailed description	<p>This variable is generated after episode splitting and shows whether overlaps exist between different sources, and if so, where. The variable is structured as a bit pattern so that every possible combination can be depicted. For this the values of the variable 'quelle' (see previous page) are simply added up, though each value per source only once. Overlaps within one source are thus not shown (see on this subject the variables 'level1' and 'nlevel1' in Section 4.4.8), apart from overlaps in the eight sub-categories within the data source for the participation-in-measures data (MTH).</p> <p>In order to determine which sources occur simultaneously, the respective highest value contained has to be deducted from the value of 'kom_quel' until the result is 0.</p> <p>Example: For one episode 'kom_quel' has the value 59:</p> <table border="1" style="margin-left: 20px;"> <tr> <td>The highest value contained in</td> <td>59</td> <td>is</td> <td>32:</td> <td><math>59 - 32 = 27</math></td> </tr> <tr> <td>The highest value contained in</td> <td>27</td> <td>is</td> <td>16:</td> <td><math>27 - 16 = 11</math></td> </tr> <tr> <td>The highest value contained in</td> <td>11</td> <td>is</td> <td>8:</td> <td><math>11 - 8 = 3</math></td> </tr> <tr> <td>The highest value contained in</td> <td>3</td> <td>is</td> <td>2:</td> <td><math>3 - 2 = 1</math></td> </tr> <tr> <td>The highest value contained in</td> <td>1</td> <td>is</td> <td>1:</td> <td><math>1 - 1 = 0</math></td> </tr> </table> <p>32, 16, 8, 2 and 1 were therefore contained in 59, i.e. in this fictitious example observations of employment, benefit receipt, FbW/TM/DSL, FF and ESF-BA would be contained for the same episode for one person.</p>	The highest value contained in	59	is	32:	$59 - 32 = 27$	The highest value contained in	27	is	16:	$27 - 16 = 11$	The highest value contained in	11	is	8:	$11 - 8 = 3$	The highest value contained in	3	is	2:	$3 - 2 = 1$	The highest value contained in	1	is	1:	$1 - 1 = 0$
The highest value contained in	59	is	32:	$59 - 32 = 27$																						
The highest value contained in	27	is	16:	$27 - 16 = 11$																						
The highest value contained in	11	is	8:	$11 - 8 = 3$																						
The highest value contained in	3	is	2:	$3 - 2 = 1$																						
The highest value contained in	1	is	1:	$1 - 1 = 0$																						

	<p>Filing in the form of a bit pattern minimises the storage space required. It can, however, easily be converted into indicator variables which show whether there is an observation from a certain source in the episode in question.</p> <p>Sample syntax in Stata:</p> <pre> generate q1      = mod(      kom_quel      , 2) ~=0 generate q2      = mod(int(kom_quel/    2) , 2) ~=0 generate q4      = mod(int(kom_quel/    4) , 2) ~=0 generate q8      = mod(int(kom_quel/    8) , 2) ~=0 generate q16     = mod(int(kom_quel/   16) , 2) ~=0 generate q32     = mod(int(kom_quel/   32) , 2) ~=0 generate q64     = mod(int(kom_quel/   64) , 2) ~=0 generate q128    = mod(int(kom_quel/  128) , 2) ~=0 generate q256    = mod(int(kom_quel/  256) , 2) ~=0 generate q512    = mod(int(kom_quel/  512) , 2) ~=0 generate q1024   = mod(int(kom_quel/1024) , 2) ~=0 mod(x,y) = x - y*int(x/y)      (modulo function) int: cuts off positions after decimal point </pre>
	1 to 2047 (possible values) / 1 to 1856 (actual values)

#### 4.4.3 Observation counter per person

See next section.

#### 4.4.4 Number of observations per person

Variable names	spell, nspell
Variable labels	observation counter per person (spell) ; number of observations per person (nspell)
Available for	all sources
Origin	generated
Anonymisation	none
Time reference	spell: variable within the individual ID and within the data record number nspell: constant within the individual ID (and thus also within the data record number)
Detailed description	<p>The observation counter '<b>spell</b>' counts all observations per person, beginning with 1. Both the original observations and those added as a result of episode splitting are counted. The data were sorted beforehand into the following order:</p> <ul style="list-style-type: none"> <li>• Individual ID (Section 4.2.1)</li> <li>• Start date of split episode (Section 4.3.3)</li> <li>• Source (Section 4.4.1)</li> </ul> <p>BeH observations are additionally sorted by:</p> <ul style="list-style-type: none"> <li>• Marginal part-time employment indicator (in ascending order, not included in the IEBS)</li> <li>• Daily wage / daily benefit rate (in descending order, Section 4.7.3)</li> </ul> <p>The variable <b>nspell</b> contains the total number of observations for the current account, i.e. the following applies: nspell = max(spell) per person.</p>
Notes	none

#### 4.4.5 Observation counter per episode

See next section.

#### 4.4.6 Number of simultaneous observations per episode

See next section.

#### 4.4.7 Observation counter per episode and source

See next section.

---

4.4.8 Number of simultaneous observations per episode and source

Variable names/ Variable labels:	Observation counter per episode: level2 No. of observations per episode: nlevel2 Observation counter per episode and source: level1 No. of observations per episode and source: nlevel1
Available for	all sources
Origin	generated
Anonymisation	none
Time reference	variable within the individual ID and the data record number
Detailed description	<p>The observation counter per episode '<b>level2</b>' numbers parallel observations of one episode from zero to n. If there are again parallel observations in another episode of the account, these are also counted starting from zero.</p> <p>The observation counter '<b>level1</b>' numbers parallel data records in one episode <u>within one data source</u> from zero to n. The sub-categories of the MTH observations (quelle = 4, 8, 16, 32, 128, 256, 512, 1024) are regarded as separate data sources. If there are parallel observations of another data source in the same episode, or there are more parallel records in another episode of the account within this source, they are each numbered from zero upwards. The observations of an account are sorted according to data source beforehand and then within the sources they are sorted into the following order:</p> <p>BeH</p> <ul style="list-style-type: none"> <li>• social security number (in ascending order, not included in the IEBS)</li> <li>• marginal part-time employment indicator (in ascending order, not included in the IEBS)</li> <li>• average daily wage (in descending order)</li> </ul> <p>LeH</p> <ul style="list-style-type: none"> <li>• social security number or client number (in ascending order, not in the IEBS)</li> <li>• type of benefit (in ascending order)</li> <li>• data record number (in ascending order)</li> </ul> <p>MTH</p> <ul style="list-style-type: none"> <li>• social security number or client number (in ascending order, not included in the IEBS)</li> <li>• data source (in ascending order)</li> <li>• data record number (in ascending order)</li> </ul> <p>BewA</p> <ul style="list-style-type: none"> <li>• social security number or client number (in ascending order, not included in the IEBS)</li> <li>• data record number (in ascending order)</li> </ul> <p>The marginal part-time employment indicator is defined as follows: employees in marginal part-time employment and short-term employment (= groups 109 and 209) are given the value 2, all other person groups are given the value 1. Sorting the data in this way means that in cases of multiple jobholding the job which is subject to social security – if there is one – is always given the value 0 for '<b>level1</b>'. Furthermore, within this sorting procedure the job with the highest remuneration is always indicated by the value 0. The BeH observation with the highest average daily wage therefore generally has the value 0 and indicates the main job. The variable for the number of observations per episode '<b>nlevel2</b>' indicates how many simultaneous observations there are within the episode, i.e. the following applies: <math>nlevel2 = \max(level2) + 1</math> per episode. The variable for the number of observations per epi-</p>

	sode and source 'nlevel1' indicates how many parallel observations there are within the episode per source, i.e. the following applies: $nlevel1 = \max(level1) + 1$ per episode and source.
Notes	Accounts with more than 2000 observations are deleted for system-inherent reasons (concerns 4 accounts) .

## 4.5 Personal status before, during and after the current observation

### 4.5.1 Employment status prior to job-search

Variable name	estatvor
Variable label	employment status prior to job-search
Available for	BewA
Origin	BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	For BewA observations this attribute shows the employment status prior to the job-search activity (regarding the job-search status see erwstat, Section 4.5.2).
Notes	From 12/2002 onwards the values of the variable were reduced substantially. The values of older data records were recoded to the currently valid values, which are thus valid for the entire period.  The details apply to the unsplit original observation and the first of the split episodes.

### 4.5.2 Employment status: person group, type of benefit, type of measure, job-search status

Variable name	erwstat
Variable label	employment status: person group, type of benefit, type of measure, job-search status
Available for	all sources
Origin	BeH (from 1999), LeH, MTH, BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	The person group variable has only been a component of the social security notifications since the introduction of the new notification procedure as of 01.01.1999. The person group can be contained in 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 variables 'school education and vocational training', 'occupational status and working hours' and 'occupation' (see Sections 4.6.6, 4.7.1 and 4.7.2) as well as other information. In many cases, however, appropriate allocations are not possible.

	<p>For <b>LeH observations</b> the variable states the grouped benefit type, i.e. details are given regarding receipt of unemployment benefit, unemployment assistance or maintenance allowance.</p> <p>For <b>MTH observations</b> the variable contains a combination of the variables 'type of measure' and 'type of benefit'. A list of all of the possible values can be found in the frequency count handbook (see <a href="http://fdz.iab.de">http://fdz.iab.de</a>).</p> <p>For <b>BewA observations</b> the variable contains the job-search status. In the BewA there is only a distinction between observations with the status "unemployed" and those with the status "jobseeker". In the IEB the additional status "incapacitated for work" is generated. Observations contain the job-search status "incapacitated for work" if</p> <ul style="list-style-type: none"> <li>• a previous observation of "unemployed" exists which joins the next observation without a gap and has "incapacitated for work" as the reason for exit (sna = 6004, see Section 4.5.4) and</li> <li>• the observation itself does not have the status "unemployed" but "jobseeker".</li> </ul>
Notes	<p>The person groups 301 to 303 in the Regulation on Data Collection and Transmission (DEÜV) are not surveyed in full.</p> <p>From time to time there are cases of individuals possessing several employment statuses simultaneously which are not legally admissible. As a result of the multitude of employment and training measures, whose requirements for participation often change over time, correcting these inconsistencies involves a great deal of time and effort. An initial list of admissible and inadmissible states can be found in Fitzenberger et al. 2005b.</p> <p>Please note: Some of the values of the variables have changed compared with the 2005 version.</p> <p>This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4)</p>

#### 4.5.3 Reason for end of observation

Variable name	grund
Variable label	reason for end of observation
Available for	BeH, LeH, BewA
Origin	BeH, LeH, BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>The reason for the end of the observation has different meanings for the different data sources:</p> <p>BeH: reason for submitting the notification  LeH: reason for end of benefit receipt  BewA: reason for exit from job-search</p> <p>In <b>BeH observations</b> the 'reason for end of observation' variable indicates the reason why the employer has reported the employment notification in question to the social security agencies. The reasons for submitting employment notifications are encoded according to the regulations of the notification procedure which has been in effect since 01.01.1999 (in accordance with DEÜV<sup>18</sup>). The reasons for submitting employment notifications are recorded in a more differentiated form than in the previous notification procedure (in accordance with the DEVO-/DÜVO regulations<sup>19</sup>). The previous reasons for submitting a notification were recoded to bring them into line with the new reasons.</p>

<sup>18</sup> DEÜV: Data Collection and Transmission Regulation (Datenerfassungs- und -übermittlungsverordnung), came into force on 1 January 1999.

	<p>The BeH and thus also the IEBS do not include all of the possible reasons for submitting a notification that may occur in the context of the notification procedure. For instance, the BeH only includes notifications that have information on wages (i.e. annual, employment interruption and end of employment notifications), but not registrations, as they contain no information on wages. However, this does not involve a loss of information as the details from a registration are transmitted again with the following annual, employment interruption or end of employment notification.</p> <p>For <b>LeH observations</b> the variable indicates the reason for the end of receipt of unemployment benefit, unemployment assistance or maintenance allowance. From 01.01.2003 some of the values of the variable have a different meaning. Which of the meanings applies depends on whether the end date of the original observation (endorig, Section 4.3.2) is smaller than or greater than / equal to 01.01.2003. The values affected by this are shown in the frequency count handbook (see <a href="http://fdz.iab.de">http://fdz.iab.de</a>).</p> <p>Reasons for the start of benefit receipt are not included in the IEBS, as this information comes from the notifications submitted by the employment agencies to the health insurance institutions regarding the beginning of a period of benefit receipt.</p> <p>In the case of <b>BewA observations</b> the variable contains the deregistration or exit reason. It must be taken into account here that the number of values of the variable was reduced from 26.04.2003 onwards. The old values are recoded to the new values as far as this is possible.</p>
Notes	<p>In the case of LeH and BewA observations it is especially necessary to take into account the changes in the meaning of the variable over time.</p> <p>The details apply to the unsplit original observation and the last of the split episodes.</p> <p>Please note: some of the values of the variables have changed compared with the 2005 version.</p> <p>This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4).</p>

#### 4.5.4 Status after exit

Variable name	sna
Variable label	status after exit
Available for	BewA, MTH
Origin	generated from the BewA variables 'reason for exit' and 'planned measure' on exit
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	This variable indicates a jobseeker's status following the end of the BewA observation. It provides information for example about whether the individual took up subsidized or non-subsidized employment after leaving unemployment. In particular in the latest available data, where parallel employment data are not yet available, this can constitute important information about the individual's whereabouts. The variable is generated from the reason for exit from the BewA (see Section 4.5.3) and the variable 'practical training (Berufspraxis) /promotion measure ID (Förder-ID) (not contained in the IEBS).

<sup>19</sup> DEVO: Data Collection Regulation (Datenerfassungsverordnung); DÜVO: Data Transmission Regulation (Datenübermittlungsverordnung); the 2nd DEVO and the 2nd DÜVO were replaced by the Data Collection and Transmission Regulation (DEÜV) on 1 January 1999.



	<p>Background:</p> <p>As a result of changes to the coArb administrative system upon which the BewA data are based, certain information which was previously recorded in the variable 'reason for deregistration' has no longer been recorded there since 26.04.2003 but is now shown instead in the variable 'planned measure' (previously the variable 'promotion measure ID') on exit. However, in the data model on which it is based the 'planned measure' is not assigned to the table on the notification status (ASU status, identifier = client number) but is a variable from the Applicant Pool Data (identifier = BewA data entry number) with its own period of validity. In order to continue to be able to assign the relevant reason for exit to each job-search record, in the research database ISAAK the 'planned measure' on exit variable was assigned to the corresponding job-search status records and a new variable was generated which combines the previous reason for deregistration and the variable 'planned measure' on exit, which has the same values before and after 2003, so there are no inconsistencies over time.</p>
Notes	<p>Some variable values are not valid across the entire period. Some of the values were dropped in 2003 as the value is no longer recorded under 'reason for exit', see the variable 'reason for end of observation' (grund). For some observations, the variable values were already changed before 2003, others were not changed until after 2003.</p> <p>The details apply to the unsplit original observation and the last of the split episodes.</p> <p>For the participants-in-measures history (MTH), however, there is only one value: 'applicant sought employment him/herself'.</p> <p>This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4). The variable is available with good data quality.</p>

## 4.6 Personal variables

### 4.6.1 Year of birth

Variable name	gebjahr
Variable label	year of birth
Available for	all sources
Origin	BeH, BA client history
Anonymisation	none
Time reference	constant within the individual ID (and thus also within the data record number)
Detailed description	<p>The year of birth is generated from the social security number if this is available.</p> <p>For accounts without a social security number, the most up-to-date year of birth is merged from the BA client history on the basis of the client number.</p>
Notes	<p>As the year is reported as a two-digit number in the social security number, the century '19' is added.</p> <p>A change of the year of birth within an account can not occur.</p> <p>The variable is generated on the basis of the date of birth.</p>

#### 4.6.2 Gender

Variable name	sex
Variable label	gender
Available for	all sources
Origin	BeH, BA client history
Anonymisation	none
Time reference	constant within the individual ID (and thus also within the data record number)
Detailed description	The gender is generated from the social security number if this is available. For accounts without a social security number, the most up-to-date gender information is merged from the BA client history on the basis of the client number.
Notes	A change of the gender within an account can not occur.

#### 4.6.3 Nationality

Variable name	nation
Variable label	nationality
Available for	all sources
Origin	BeH, LeH, BA client history
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	The variable 'nationality' is available in the data in a slightly aggregated form.
Notes	This variable is not corrected, i.e. different nationalities may occur in one account for a point in time or over time, which are based on different details in the source data and are not necessarily associated with an actual change of nationality. The definition of nationalities refers to the time when the data were collected, so countries that no longer exist (e.g. the Soviet Union or Yugoslavia) are included. This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4).

#### 4.6.4 Severe disability status

Variable name	schweb
Variable label	severe disability status
Available for	BewA, MTH
Origin	BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	This variable comes from the BewA. The severe disability status which is valid on the day before the start of participation in an employment or training measure ( <b>MTH observations</b> ) or the start of job-search activity ( <b>BewA observations</b> ) is merged from the BewA.

Notes	<p>This information is required in order to determine whether an individual has a disability in accordance with the Severely Disabled Persons Act (Schwerbehindertengesetz). A disability is regarded as severe from a 50% degree of disability.</p> <p>This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4). The variable is available with good data quality.</p>
-------	--

#### 4.6.5 School-leaving qualification

Variable name	schbild
Variable label	school-leaving qualification
Available for	BewA, MTH
Origin	BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>This variable comes from the Applicant Pool Data (BewA) and contains the school-leaving qualification for the MTH and BewA data sources. (<b>MTH observations:</b> valid from the day before the start of the employment or training measure, <b>BewA observations:</b> valid from the start date of unemployment or the job-search status)</p>
Notes	<p>Details on the school-leaving qualifications of employees are contained in the variable 'school education and vocational training'. See also Section 4.6.6 on this subject.</p> <p>The details in the MTH and BewA data records appear to be more reliable than those in the Employee History File (BeH). The reason for this is that the BeH data come from the employment notification procedure, where the school-leaving qualification does not affect any social security entitlements and is therefore presumably reported with less care.</p> <p>This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4). The variable is available with good data quality.</p>

#### 4.6.6 School education and vocational training

Variable name	bild
Variable label	school education and vocational training
Available for	BeH, MTH, BewA
Origin	BeH, BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number

Detailed description	<p>For <b>BeH observations</b> this variable contains the vocational training code (B2 code) from the employment statistics, i.e. the details are taken from the 'employment details' reported in the employment notification procedure. The indicator 'school education and vocational training' is a combined variable which is used to gather information on both the employee's school and vocational qualifications (the highest qualification gained in each case). Here the school qualifications are to be established first and then the vocational qualifications. With regard to the school qualification the variable can only be interpreted as a dichotomous variable: the individual either possesses an upper secondary school-leaving certificate or s/he does not. If the individual does not have this school-leaving certificate, it is not possible to determine with certainty from the data whether s/he actually gained a qualification at other schools or whether only attendance at such a school is recorded. With regard to vocational training, in the case of the values 5 (degree from a university of applied sciences) and 6 (university degree) no other vocational qualifications are determined. Further vocational training (e.g. courses at language schools and schools for interpreters) and special skills which were acquired by means of practical activities are not counted as vocational training here.</p> <p>For <b>MTH and BewA observations</b> the data on the vocational qualification gained come from the BewA. For MTH observations the information is merged at the start of the measure, for BewA observations at the start of the job-search.</p>
Notes	<p>For certain sub-groups there is a large proportion of missing details in the BeH observations because this variable is not of importance as regards social security. Changes in the vocational training status frequently occur at the same time as a change of establishment. This is because the notification data are compiled anew in the new firm. If, for example, an employee has gained a higher qualification via a part-time further training course while still working then this change of status is probably not recorded until s/he joins a new firm. It can generally be assumed that when a person is employed in a firm for a longer period, the personal data that they reported at the beginning is simply continued.</p> <p>Consistency checks on the education variables ('school-leaving qualification', Section 0 and 'school education and vocational training', Section 4.6.6) conducted by Fitzenberger et al. (2005b, S. 36 ff.) reveal a not inconsiderable number of inconsistencies, both between the data sources and over time. The information in the MTH and the BewA data records appears to be more reliable than that in the BeH.</p> <p>Fitzenberger et al. (2005a) make concrete suggestions for correcting the education variable BILD in the IABS, which corresponds to the IEBS variables of the BeH observations. In the case of the IEBS it is advisable to use the information in the MTH and BewA observations in addition.</p> <p>New frequency counts reveal a substantial increase in the number of missing values for younger entry cohorts (up to 30%).</p> <p>This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4). The variable is currently not available to an adequate extent for data from VerBIS and can therefore not be analysed.</p>

#### 4.6.7 Skills level

Variable name	quali
Variable label	skills level
Available for	BewA
Origin	BewA
Anonymisation	none
Time reference	variable within the individual ID and the data record number

Detailed description	This variable takes into account not only vocational training but also occupational experience. In the case of equal competence as a result of educational background, performance or experience, a skills level can also be entered when the applicant does not hold a corresponding qualification.
Notes	This variable contains the skills levels of job applicants as they are recorded in coArb. The variable is no longer supplied from VerBIS. This variable is affected by the changeover to the VerBIS procedure (see Section 4.5.4).

#### 4.6.8 Marital status

Variable name	famstand
Variable label	marital status
Available for	BeH and MTH
Origin	BeH and MTH
Anonymisation	none
Time reference	variable within the individual ID and the data record number
Detailed description	<p>This variable describes the applicants' marital status (BewA). Marital status includes categories of living in a joint household with children or a partner. However, it is not possible to distinguish clearly whether</p> <ul style="list-style-type: none"> <li>• a married person / person living in a joint household is bringing up children,</li> <li>• a married person / person living in a joint household / person living alone has any children at all,</li> <li>• a lone parent / person living in a joint household / person living alone is married</li> </ul> <p>Variable values smaller than 3 are values from the benefit receipt data (LeH), which only include two values. No comparison of the information from the two data sources was made.</p>
Notes	This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4). The variable is available with good data quality. The information from the LeH is of poorer quality.

#### 4.6.9 Number of children in the household

Variable name	kind
Variable label	number of children in the household
Available for	BewA
Anonymisation	none
Origin	BewA
Time reference	variable within the individual ID and the data record number
Detailed description	<p>This variable indicates the number of children aged under 15 living in the household.</p> <p>The reason for collecting this information is that the existence of children requiring childcare is of importance for assessing the suitability of job offers. Benefit recipients with children under the age of three can not reasonably be expected to exercise an occupation, see. §10 (1) No. 3 Social Code Book II (SGB II).</p>

Notes	<p>It is uncertain whether only children actually under the age of 15 are counted.</p> <p>The possible number of children changed with the switchover to VerBIS:</p> <ul style="list-style-type: none"> <li>• CoArb: it was not possible to record a number of children &gt; 10</li> <li>• VerBIS: up to 99 children can be recorded</li> </ul> <p>In the IEBS dataset, cases with more than three children were combined into one class.</p> <p>This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4). The variable is available with good data quality.</p>
-------	---

#### 4.6.10 Health impediments

Variable name	gesund_ein
Variable label	health impediments
Available for	BewA
Origin	BewA
Anonymisation	none
Time reference	variable within the individual ID and the data record number
Detailed description	Applicants' existing health problems
Notes	<p>Details on the consequences that the health problems have for job placement are only available in the Applicant Pool Data (BewA). The variable only provides information about whether any health problems exist or not.</p> <p>This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4). The variable is available with good data quality.</p>

## 4.7 Data on the employment relationship and job-search activity

### 4.7.1 Occupational status and working hours

Variable name	stib
Variable label	Occupational status and working hours
Available for	BeH
Origin	BeH
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>The variable 'occupational status and working hours' originates from the "employment details" submitted by the employer in the context of the employment notification procedure.</p> <p>The category "employees in vocational training" (value 0) covers not only trainees/apprentices, placement workers and interns but also semi-skilled trainees, students at colleges for health occupations and participants in subsidized further vocational training, retraining and induction training.</p> <p>Full-time and part-time employees are differentiated according to the ratio between the contracted hours and the usual working hours in the establishment. Part-time employees (values 8 and 9) are divided into two sub-groups in which the dividing line is half of the usual working hours of a full-time employee. Both for part-time employees and for master craftsmen and foremen it is only possible to assign the status of a blue-collar or a white-collar worker via the respective pension insurance institution.</p>
Notes	<p>Owing to the way that the data are recorded in the employment notification procedure, the occupational status is only available for full-time employees. For part-time employees, on the other hand, only the status as a part-time employee is known and not the occupational status.</p>

## 4.7.2 Occupation

Variable name	beruf
Variable label	Occupation
Available for	all sources
Origin	BeH, BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>The values of the occupations correspond to the occupational classification of the Federal Employment Services: "Classification of Occupations. Systematic and Alphabetical Directory of Job Titles". This contains approx. 25,000 job titles; the aggregation level of the occupational classification consists of a 3-digit code comprising about 330 values.</p> <p>In the case of <b>BeH observations</b> the occupational title of the job currently performed by the employee comes from the 'employment details' submitted by the employer. For this the employer encodes the employee's job on the basis of the occupational classification (3-digit code). 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.</p> <p>What must be borne in mind with all of the employee history variables is that certain groups of employees are not subject to social security contributions. As a consequence, the distribution of occupations is not representative of the total population of people in gainful employment.</p> <p><b>MTH</b> and <b>BewA observations</b> contain the last occupation held as taken from the Applicant Pool Data (BewA). <b>MTH observations</b> contain this information from the BewA referring to the day prior to the start of the measure, in the <b>BewA observations</b> the information refers to the status at the beginning of the spell of job-search activity.</p>
Notes	<p>The large proportion of missing values in the BewA observations can be partly explained by the fact that many jobseekers have never been in employment before or have not been employed for some time (see variable 'employment status prior to job-search', Section 4.5.1).</p> <p>This variable is affected by the changeover to the Verbis procedure (see Section 4.5.4). The variable is available with good data quality.</p>

## 4.7.3 Daily wage / daily benefit rate

Variable name	tentgelt,
Variable label	daily wage / daily benefit rate
Available for	BeH, LeH
Origin	BeH, LeH
Anonymisation	none (tentgelt_d: aggregated)
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>In <b>BeH observations</b> this variable shows the employee's gross daily wage. It was calculated from the fixed-period wages reported by the employer and the duration of the unsplit original notification period in calendar days.</p> <p>In cases where an error arises in the wage, the daily wage is deleted. The observation is retained, however.</p>



	<p>In the period 1975 to 1998 employers in principle only reported the earnings which were subject to social security contributions, in other words earnings above the marginal part-time income threshold and up to the upper earnings limit for statutory pension insurance. After the inclusion of marginal part-time employees in the employment notification procedure as of 01.04.1999, earnings below the marginal part-time income threshold are also reported. The upper earnings limit still applies as the upper ceiling. In exceptional cases earnings notifications may nonetheless exceed the upper earnings limit, however. Generally this can probably be attributed to the payment of annual bonuses which the employer can add to the regular earnings in the annual, employment interruption or end of employment notifications. In this case it is irrelevant whether the upper earnings limit in the statutory pension insurance which is decisive for the notification period is exceeded as a result of this addition. However, such earnings notifications could also be due to incorrect details in the employment period or the amount of wages paid, though this is likely to be seldom the case given the relevance of this information for social security contributions. The marginal part-time income threshold and the upper earnings limit for statutory pension insurance differ from year to year, they also vary depending on the pension insurance institution as well as between eastern and western Germany. The decisive factor here is the location of the establishment. An overview of these limits and thresholds is available from the FDZ.</p> <p>For <b>LeH observations</b> the variable shows the daily benefit rate of unemployment benefit, unemployment assistance or maintenance allowance. It must be taken into account here that for observations with an original start date prior to 01.01.1998 the daily benefit rate applies to working days, while for observations with an original start date from 01.01.1998 onwards it applies to calendar days.</p>
Notes	<p>Daily wages and benefit rates are shown in Euros.</p> <p>Daily wages equal to or close to 0 can occur in cases of employment interruption notifications or for legally existing employment relationships without pay – so-called 'dormant employment relationships' (daily wages greater than 0 or close to 0 would be possible for example if special distributions of funds are made which are relevant for income calculations). A typical example of a dormant employment relationship concerns women on parental leave.</p>

#### 4.7.4 Transition zone

Variable name	gleitz
Variable label	transition zone
Available for	BeH, since 01.04.2003
Origin	BeH
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	This variable indicates whether the employment notification relates to employment in the low-wage sector, within the so-called transition zone. Jobs in the transition zone have a gross wage of € 400 to € 800 for which the employee only has to pay a reduced social security contribution. The corresponding legislation has been in force since 01.04.2003

#### 4.7.5 Classification of economic activities (73, 93, 03)

Variable name	w73, w93, w03
Variable label	Classification of economic activity 73; classification of economic activity 93, classification of economic activity 03
Available for	BeH, since 1999
Origin	BeH

Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>This variable indicates the economic activity as a 3-digit code in accordance with the classification WZXX, i.e. the economic group. WZXX stands for the "Classification of Economic Activities for the Statistics of the Federal Employment Services" ("Klassifikation der Wirtschaftszweige für die Statistik der Bundesanstalt für Arbeit"), editions 1993, 1973 and 2003". The economic activity according to the 1993 classification has only been included since 1999 (since the introduction of the new notification procedure) and supersedes the WZ73 classification system. In 2003 another new classification system (WZ03) was introduced.</p> <p>The classification of economic activities WZXX is used to record the economic activities of statistical or local units, i.e. "establishments"<sup>20</sup>. A local unit is to be understood as a firm or part of a firm<sup>21</sup> located in a specifically defined location<sup>22</sup>. For a more detailed account see for example. <a href="http://www.statistik-portal.de/Statistik-Portal/klassiWZ93.asp">http://www.statistik-portal.de/Statistik-Portal/klassiWZ93.asp</a> and <a href="http://www.statistik-portal.de/Statistik-Portal/de_klassiWZ03.asp">http://www.statistik-portal.de/Statistik-Portal/de_klassiWZ03.asp</a>.</p> <p>As an example of WZ93: this is based on the Statistical Classification of Economic Activities in the European Community (NACE Rev.1)<sup>23</sup>, which has four structural levels, the first two of which are in turn based on the international standard ISIC Rev.3<sup>24</sup>. In order to guarantee comparability of the statistical data in the different member states, this basic structure should be adopted in its entirety. The inclusion of all the systems can be used, if applicable, to generate classifications which are consistent over time for the analyses. Alternatively, old classifications can also be continued over time via the establishment indicator.</p> <p>Each establishment may only be assigned one code. If an establishment is active in different economic sectors, the main economic activity must be determined. This is based on the gross value added at factor cost or on suitable substitute values (e.g. number of employees). In its practical work the BA has to be content with the self-assessment of the establishments for the classification.</p>
Notes	The economic activity is merged via the establishment number and is missing if this number is not valid.

#### 4.7.6 Willingness to seek employment throughout Germany

Variable name	mobil
Variable label	willingness to seek employment throughout Germany
Available for	BewA
Origin	BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number

<sup>20</sup> The statistical or local unit used at the BA is in principle the DEVO/DÜVO establishment defined in the "Establishment Data Procedure – co-Ber-coStat".

<sup>21</sup> Workshop, plant, shop, office, mine, warehouse etc.

<sup>22</sup> The concept of a specifically defined location is very restricted: two production units belonging to the same firm but located in different places are to be regarded as two local units, even if the two locations are in the same municipality or community.

<sup>23</sup> NACE Rev.1: "Nomenclature générale des activités économiques dans les communautés européennes", first revision (1990); the first version was published in 1970.

<sup>24</sup> ISIC Rev.3: "International Standard Industrial Classification of All Economic Activities", third revision (1990).

Detailed description	This variable describes a person's potential regional mobility. When registering job-search the jobseeker has to indicate whether s/he is seeking work anywhere in Germany or whether s/he is restricted to certain regions. If only selected regions are indicated, the individual is then asked whether other regions could nevertheless be considered. If this is the case, a person is regarded as "regionally mobile", and otherwise they are not. If a person indicates that they are seeking work across Germany s/he is automatically regarded as "mobile".
Notes	none

#### 4.7.7 BA client group

Variable name	kunden_gr
Variable label	BA client group
Available for	BewA
Origin	BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	This variable contains the classification of the BA clients into different types, such as market client, client to be activated, client to be supported and counselling client. This differentiation takes into account the need for action with regard to the individual's obstacles to integration.
Notes	none

#### 4.7.8 Type of termination

Variable name	art_kuend
Variable label	type of termination of last job
Available for	BewA
Origin	BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	This variable describes how the last employment or training relationship was terminated before a new period of job-search. It can therefore be used to identify job-to-job placements.
Notes	none

## 4.7.9 Desired working hours of the job sought

Variable name	arbzeit
Variable label	desired working hours of the job sought
Available for	BewA
Origin	BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	During the placement procedure jobseekers indicate how many working hours the job they are seeking should have.
Notes	VerBIS effects are discernible in this variable as this information is not obligatory in VerBIS. The degree of completeness is to be improved, however, by means of the relevant work instructions.

## 4.7.10 Duration of remaining entitlement to unemployment benefit

Variable name	restanspruch
Variable label	duration of remaining entitlement to unemployment benefit
Available for	LeH
Origin	LeH
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>The duration of remaining entitlement to unemployment benefit payments is shown in the benefit recipient history (LeH). It refers to the end date of a period of benefit receipt.</p> <p>Remaining entitlement is the entitlement that remains after the end of the current period of benefit receipt. Remaining entitlement continues for 7 years.</p> <p>The statement "missing" can be assumed to express a remaining entitlement of 0 days.</p>
Notes	If the end date is before 01.01.1998, the remaining entitlement is reported in working days, after this date it is reported in calendar days. This entitlement counts up to a maximum of the longest possible duration for the individual's respective age.

## 4.7.11 Planned start date

Variable name	begplan
Variable label	planned start date
Available for	MTH
Origin	MTH
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	Planned start date of a measure. This variable is only available for datasets regarding employment and training measures. The planned start date can, but need not necessarily, be identical to the start date of the original period of validity.
Notes	none

#### 4.7.12 Planned end date

Variable name	endplan
Variable label	planned end date
Available for	MTH
Origin	MTH
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	Planned end date of a measure. This variable is only available for datasets regarding employment and training measures. The planned end date can, but need not necessarily, be identical to the end date of the original period of validity.
Notes	none

### 4.8 Data on the employing establishment

#### 4.8.1 Number of regular employees

Variable name	az_hpt
Variable label	number of regular employees in the establishment
Available for	BeH
Origin	BHP, Establishment History Panel
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	This variable indicates the number of employees whose main job was in the establishment. The main job is defined as the job with the highest daily wage rate. If there is more than one employment notification with the same wage rate as of 30 June of a year for one person, the job with the longest employment duration counts. Marginal part-time work is only counted as the main job if the individual does not also have a job which is subject to social security contributions at the same time.
Notes	This information is generated. It is not data from the employment notification procedure.

## 4.8.2 Year when establishment was founded

Variable name	grd_dat
Variable label	year when establishment was founded
Available for	BeH
Origin	BHP, Establishment History Panel
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	Based on the first date when the establishment appears in the Employee and Benefit Recipient History (BLH).
Notes	This information is generated. It is not data from the employment notification procedure.

## 4.9 Location data

### 4.9.1 Place of work: federal state (Bundesland), district (Kreis)

Variable name	ao_bula (ao_kreis)
Variable label	place of work: federal state (Bundesland); also place of work: district (Kreis)
Available for	BeH
Origin	BeH
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>This variable contains the location of the establishment in which the person is employed. The coding is based on an 8-digit local authority code (Gemeindeschlüssel) developed by the Federal Statistical Office (Statistisches Bundesamt). The first two digits of this code indicate the federal state (Bundesland), positions 1-3 indicate the regional authority (Regierungsbezirk), positions 1-5 show the district authority (Kreis) and positions 1-8 show the municipality or local authority (Gemeinde). Federal states without a regional authority have a 0 in the third position. Until 1999 the place of work was encoded at the level of municipality or local authority.</p> <p>As the district boundaries change regularly over time, cases occur in which the local authority code changes without the establishment relocating, if the territorial allocations of the districts are not updated. In order to guarantee consistent regional allocations across the entire observation period, the data in the BeH observations were recoded to the territorial allocation of 31.12.2005, i.e. in all calendar years an establishment location is assigned to a municipality in accordance with the boundaries as of 31.12.2005.</p>
Notes	<p>Owing to its particular sensitivity with regard to data protection legislation, the variable of place of work at district level (ao_kreis) is only made available on application and only in well-founded cases (→ sensitive variables, see Section 3.7). Otherwise the data contain only the federal state (ao_bula).</p> <p>In 1991 missing values occur increasingly often. This can be put down to the fact that the BeH for 1991 contains incorrect local authority codes for the federal states of eastern Germany, which were recoded to a missing value.</p> <p>The place of work is merged via the establishment number and is missing if this number is not valid.</p>

### 4.9.2 Place of residence: federal state (Bundesland), district (Kreis)

Variable name	wo_bula (wo_kreis)
Variable label	place of residence: federal state (Bundesland); (also place of residence: district (Kreis)
Available for	BeH (from 1999), MTH, BewA
Origin	BeH, MTH, BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number

Detailed description	<p>This variable contains the person's place of residence. The coding is based on an 8-digit local authority code (Gemeindeschlüssel) developed by the Federal Statistical Office (Statistisches Bundesamt). The first two digits of this code indicate the federal state (Bundesland), positions 1-3 indicate the regional authority (Regierungsbezirk), positions 1-5 show the district authority (Kreis) and positions 1-8 show the municipality or local authority (Gemeinde). Federal states without a regional authority have a 0 in the third position.</p> <p>As the district boundaries change regularly over time, cases occur in which the local authority code changes without the individual having relocated, if the territorial allocations of the districts are not updated. In order to guarantee consistent regional allocations across the entire observation period, the data in the <b>BeH observations</b> were recoded to the territorial allocation of 31.12.2005, i.e. in all calendar years a place of residence is assigned to a municipality in accordance with the boundaries as of 31.12.2005.</p> <p>For <b>MTH</b> and <b>BewA observations</b> the territorial allocations were not updated. As a result of reforms in the territorial allocations of districts in Saxony-Anhalt and Thuringia this leads to newly designated districts and urban districts (kreisfreie Stadt) in 2007 (coding numbers 14000 to 14099 and 15001 to 15091). The allocation to employment agency district and federal state is unaffected by this.</p>
Notes	<p>Owing to its particular sensitivity with regard to data protection legislation, the variable of place of residence at district level (wo_kreis) is only made available on application and only in well-founded cases (→ sensitive variables, see Section 3.7). Otherwise the data contain only the federal state (wo_bula).</p>

#### 4.9.3 Place of work: regional directorate (employment agency area)

Variable name	ao_rd (ao_aa)
Variable label	place of work: regional directorate (employment agency area)
Available for	BeH, MTH, BewA
Origin	BeH, MTH, BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>The Federal Employment Agency is divided into:</p> <ul style="list-style-type: none"> <li>• regional directorates (regional employment offices until 31.12.2003),</li> <li>• employment agencies (Agenturen für Arbeit) (employment offices until 31.12.2003),</li> <li>• local employment offices.</li> </ul> <p>This variable contains</p> <ul style="list-style-type: none"> <li>• the local employment office of the place of work for <b>BeH observations</b>,</li> <li>• the local employment office of the place where the measure is conducted in <b>MTH observations</b> and</li> <li>• the statistical employment office for <b>BewA observations</b>, i.e. the one which has the records of the Applicant Pool Data.</li> </ul> <p>As the boundaries of the employment office areas have changed over time, cases occur in which the employment office code changes without the firm actually relocating to a different district, if the territorial allocations of the districts are not updated. In order to guarantee consistent regional allocations across the entire observation period, the data in the <b>BeH observations</b> were recoded to the territorial allocation of 31.12.2007, i.e. in all calendar years an observation is assigned to an employment office area in accordance with the boundaries as of 31.12.2007.</p>



	For <b>MTH</b> and <b>BewA observations</b> the territorial allocations were not updated, i.e. the data on the employment offices corresponds to the status at the beginning of the original observation.
Notes	Owing to their particular sensitivity with regard to data protection legislation, location data at employment agency level (ao_aa) are only made available on application and only in well-founded cases (→ sensitive variables, see Section 3.7). Otherwise the data contain only the regional directorate (ao_rd).  The place of work of the <b>BeH observations</b> is merged via the establishment number and is missing if this number is not valid.

#### 4.9.4 Place of residence: regional directorate (employment agency area)

Variable name	wo_rd (wo_aa)
Variable label	place of residence: regional directorate (employment agency area)
Available for	BeH (from 1999), LeH, MTH, BewA
Origin	BeH, LeH, BewA
Anonymisation	none
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>The Federal Employment Agency is currently divided into:</p> <ul style="list-style-type: none"> <li>• regional directorates (regional employment offices until 31.12.2003),</li> <li>• employment agencies (Agenturen für Arbeit) (employment offices until 31.12.2003),</li> <li>• local employment offices.</li> </ul> <p>This variable contains</p> <ul style="list-style-type: none"> <li>• the local employment office of the place of residence for <b>BeH observations</b> (available from 1999 onwards),</li> <li>• the <b>employment agency</b> responsible for the benefit payments in <b>LeH observations</b></li> <li>• the local employment office of the place of residence taken from the BewA on the day prior to the start of the measure in <b>MTH observations</b> and</li> <li>• the local employment office of the place of residence at the beginning of the job-search period for <b>BewA observations</b>.</li> </ul> <p>As the boundaries of the employment office areas have changed over time, cases occur in which the employment office code changes without the individual actually relocating to a different district, if the territorial allocations of the districts are not updated. In order to guarantee consistent regional allocations across the entire observation period, the data in the <b>BeH observations</b> were recoded to the territorial allocation of 31.12.2007, i.e. in all calendar years an observation is assigned to an employment office area in accordance with the boundaries as of 31.12.2007.</p> <p>For <b>MTH</b> and <b>BewA observations</b> the territorial allocations were not updated, i.e. the data on the employment offices corresponds to the status at the beginning of the original observation.</p>
Notes	Owing to their particular sensitivity with regard to data protection legislation, location data at employment agency level (wo_aa) are only made available on application and only in well-founded cases (→ sensitive variables, see Section 3.7). Otherwise the data contain only the regional directorate (wo_rd).

## 4.9.5 Place of residence: employment agency region type

Variable name	wo_aatyp06
Variable label	place of residence: employment agency region type
Available for	BeH (from 1999), LeH, MTH, BewA
Anonymisation	none
Origin	BeH, LeH, BewA
Time reference	variable within the individual ID, constant within the data record number
Detailed description	<p>The employment agencies are assigned to a certain labour market type.</p> <p>The region indicator is used to classify the regions with regard to the labour market situation prevailing there in the year 2006.</p> <p>This variable contains the classification of employment agency areas according to the labour market situation prevailing there, which was developed by Blien et al. (2004). The 12 comparison types can be aggregated to 5 strategy types. The variable is allocated with reference to the place of residence (see 'place of residence: regional directorate (employment agency area', Section 4.9.4).</p>
Notes	none

## 4.9.6 Place of residence abroad

Variable name	wo_ausl
Variable label	place of residence abroad
Available for	BeH (from 1999), LeH, MTH, BewA
Anonymisation	none
Origin	BeH, LeH, BewA
Time reference	variable within the individual ID, constant within the case ID
Detailed description	<p>This variable is based on the information about residence abroad in the variable 'place of residence' (wo_kreis). A recoding procedure makes it possible identify places of residence abroad. The variable distinguishes between: Europe (in detail), Africa, America, Asia, Oceania, elsewhere abroad.</p>
Notes	<p>Owing to its particular sensitivity with regard to data protection legislation, the variable 'residence abroad' is only made available on application and only in well-founded cases (→ sensitive variables, see Section 3.7).</p>

## 5 References

Blien, Uwe, Franziska Hirschenauer, Manfred Arendt, Hans Jürgen Braun, Dieter-Michael Gunst, Sibel Kilcioglu, Helmut Kleinschmidt, Martina Musati, Hermann Roß, Dieter Vollkommer und Jochen Wein (2004), Typisierung von Bezirken der Agenturen für Arbeit, Zeitschrift für Arbeitsmarktforschung, Jg. 37, H. 2. S. 146-175.

Blien, Uwe; Hirschenauer, Franziska; Arendt, Manfred; Braun, Hans Jürgen; Gunst, Dieter-Michael; Kilcioglu, Sibel; Kleinschmidt, Helmut; Musati, Martina; Roß, Hermann; Vollkommer, Dieter; Wein, Jochen (2004): Typisierung von Bezirken der Agenturen für Arbeit. In: Zeitschrift für ArbeitsmarktForschung, Jg. 37, H. 2, S. 146-175

Engelhard, Astrid, Oberschachtsiek, Dirk und Scioch, Patrycja (2008), Datengenese zweier Datenkonzepte – MTG (Maßnahme-Teilnahme-Grunddatei) und ISAAK (Instrumente Aktiver Arbeitsmarktpolitik). Eine Betrachtung ausgewählter Fälle am Beispiel der Förderung im Rahmen des ESF-BA-Programms. FDZ Methodenreport 08/2008

Fitzenberger, Bernd und Ralf Wilke (2004), Unemployment Durations in West-Germany Before and After the Reform of the Unemployment Compensation System During the 1980s , ZEW Discussion Paper No. 04-24, Mannheim.

Fitzenberger, Bernd, Aderonke Osikominu und Robert Völter (2005a), Imputation Rules to Improve the Education Variable in the IAB Employment Subsample, ZEW Discussion Paper No. 05-10.

Fitzenberger, Bernd, Michael Lechner, Ruth Miquel, Stephan Wiehler, Conny Wunsch, Martin Biewen, Sonja Lischke, Aderonke Osikominu, Tobias Wenzel (2005b), Die Beschäftigungswirkungen der FbW-Maßnahmen 2000-2002 auf individueller Ebene: Eine Evaluation auf Basis der prozessproduzierten Daten des IAB: Eine Evaluation auf Basis der prozessproduzierten Daten des IAB – vorläufiger, unveröffentlichter Zwischenbericht-.

Jaenichen, Ursula, Thomas Kruppe, Gesine Stephan, Britta Ullrich und Frank Wießner (2005), You can split it if you really want: Korrekturvorschläge für ausgewählte Inkonsistenzen in IEB und MTH, FDZ-Datenreport 4/2005.

Lee, Sokbae und Ralf Wilke (2005), Reform of Unemployment Compensation in Germany: A Non-parametric Bounds Analysis Using Register Data, ZEW Discussion Paper No. 05-29, Mannheim.

Ratgeber zur Sozialversicherung 2005, Broschüre des Deutsche BKK Arbeitgeberservice, [http://www1.deutschebkk.de/Service/Arbeitgeberservice/Wissen\\_von\\_A-Z/5B2D7F649D116FBDC1256DC900510B3E/RatgeberSV\\_2009\\_Internet.pdf](http://www1.deutschebkk.de/Service/Arbeitgeberservice/Wissen_von_A-Z/5B2D7F649D116FBDC1256DC900510B3E/RatgeberSV_2009_Internet.pdf) , S. 32 (Link vom 04.05.2009)

## 6 Appendix

### 6.1 List of abbreviations

AA	Agentur für Arbeit/ Arbeitsamt	employment agency / employment office
AB	Arbeitsberatung	careers advice
ABM	Arbeitsbeschaffungsmaßnahmen	job-creation measures
AEZ	Arbeitsentgeltzuschuss	wage subsidies for the release of unskilled workers for further training
AG	Arbeitgeber	employer
ALG	Arbeitslosengeld	unemployment benefit
ALHI	Arbeitslosenhilfe	unemployment assistance
ASU	Arbeitsuche	job-search
AU	Arbeitsunfähigkeit	incapacity for work
AV	Arbeitsvermittlung	job placement (services)
BA	Bundesagentur für Arbeit	Federal Employment Agency
BeH	Beschäftigten-Historik	Employee History
BewA	Bewerberangebot	Applicant Pool Data
BHD	Beschäftigten-Historik-Datei	Employment History File
BHI	Beschäftigungshilfe für Langzeitarbeitslose	job opportunities for the long-term unemployed
BLH	Beschäftigten- und Leistungsempfänger-Historik	Employee and Benefit Recipient History
BSI	Beschäftigung schaffende Infrastrukturförderung	development of job-creating infrastructural measures
coArb	Computerunterstützte Arbeitsvermittlung (operatives Verfahren zur Verwaltung der Vermittlung (Altverfahren))	computer-aided job placement (procedure for the administration of job placements – old procedure)
coLei	Computerunterstützte Leistungsgewährung (operatives Verfahren zur Leistungsgewährung (Altverfahren))	computer-aided benefit administration (procedure for the administration of benefit payments – old procedure)
coLibri	Operatives Verfahren zur Leistungsgewährung (Neuverfahren)	computer-aided benefit calculation and information system (new procedure)
coSach	Operatives Verfahren zur Verwaltung von Maßnahmen und Teilnahmen	computer-aided procedure for the administration of incentive measures and participation in measures
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
DSL	Deutschsprachlehrgang	German language course
DÜVO	Zweite VO über die Datenübermittlung auf maschinell verwertbaren Datenträgern im Bereich der Sozialversicherung und der BA – Datenübermittlungs-Verordnung –	Data Transmission Regulation - second regulation on the transfer of data on machine-readable data media in the field of social security and the BA
DWH	Data Warehouse	Data Warehouse

EGHI	Eingliederungshilfe	integration assistance
EGN	Eingliederungszuschuss bei Neugründung	recruitment subsidies for new businesses
EGZ	Eingliederungszuschuss	settling-in allowance
ESF	Europäischer Sozialfonds	European Social Fund
EV	Eingliederungsvertrag	integration agreement
ExGZ	Existenzgründerzuschuss	business start-up allowance
EZV	Einstellungszuschüsse bei Vertretung	recruitment subsidies for (employees providing) temporary cover
FbW	Förderung beruflicher Weiterbildung	promotion of further vocational training
FDZ	Forschungsdatenzentrum	Research Data Centre
FELEG	Gesetz zur Förderung der Einstellung der landwirtschaftlichen Erwerbstätigkeit	German law on the promotion of discontinuance of farming activities
FF	Freie Förderung	independent employment promotion measures
HOGA	Hotel- und Gaststättenvermittlung	job placement service for the hotel and restaurant sector
IAB	Institut für Arbeitsmarkt- und Berufsforschung	Institute for Employment Research
IABS	IAB-Beschäftigtenstichprobe	IAB Employment Samples
KuG	Kurzarbeitergeld	short-time working allowance
LeH	Leistungsempfänger-Historik	Benefit Recipient History
LIAB	Linked-Employer-Employee-Daten des IAB	Linked Employer-Employee Data from the IAB
LKZ	Lohnkostenzuschuss	wage subsidy
LZA	Langzeit-Arbeitslosigkeit	long-term unemployment
MTH	Maßnahme-Teilnehmer-Historik	Participants-in-Measures History file
NACE	Allgemeine Systematik der Wirtschaftszweige in den Europäischen Gemeinschaften des Statistischen Amtes der Europäischen Gemeinschaften	Statistical Classification of Economic Activities in the European Community (Nomenclature générale des activités économiques dans les communautés européennes)
PSA	Personal-Service-Agentur	Personnel Services Agency
SAM	Strukturanpassungsmaßnahmen	structural adjustment measures
Schwbg	Gesetz zur Sicherung der Eingliederung Schwerbehinderter in Arbeit, Beruf und Gesellschaft – Schwerbehindertengesetz-	law to guarantee the integration of persons with severe disabilities into employment and society – Severely Disabled Persons Act
SGB	Sozialgesetzbuch	Social Code
TM	Trainingsmaßnahme	short-term training schemes
UEG	Überbrückungsgeld	bridging allowance
UHG	Unterhaltsgeld	maintenance allowance
VDR	Verband deutscher Rentenversicherungsträger	Association of German Pension Funds
ZND	Zahlungsnachweisdatei	Benefit Payments Control File
zPDV	Zentrale Personen-Daten-Verwaltung	Central Administration of Personal Data

## 6.2 Alphabetical list of variables

Variable name	Page	Description
persnr		Individual ID
satznr		Data record number
betnr		Establishment number
begorig		Original start date of observation
endorig		Original end date of observation
begepi		Start date of split episode
endeipi		End date of split episode
quelle		Source of the information
kom_quel		Source combination
spell		Observation counter per person
nspell		Number of observations per person
level2		Observation counter per episode
nlevel2		Number of observations per episode
level1		Observation counter per episode and source
nlevel1		Number of observations per episode and source
estatvor		Employment status prior to job-search
erwstat		Employment status: person group, type of benefit, type of measure, job-search status
grund		Reason for end of observation
sna		Status after exit
gebjahr		Year of birth
sex		Gender
nation		Nationality
schweb		Severe disability status
schbild		School-leaving qualification
bild		School education and vocational training
quali		Skills level
famstand		Marital status
kind		Number of children in the household
gesund_ein		Health impediments
stib		Occupational status and working hours
beruf		Occupation
tentgelt		Daily wage / daily benefit rate
gleitz		Transition zone
w73		Economic Activity 73
w93		Economic Activity 93
w03		Economic Activity 03
mobil		Willingness to seek employment throughout Germany
kunden_gr		BA client group

art_kuend		Type of termination of last job
arbzeit		Desired working hours of the job sought
restanspruch		Duration of remaining entitlement to unemployment benefit
endplan		Planned end date of a measure
begplan		Planned start date of a measure
az_hpt		Number of regular employees in the establishment
grd_jahr		Year when employing establishment was founded
ao_bula		Place of work: federal state (Bundesland); district (Kreis)
wo_bula		Place of residence: federal state (Bundesland); district (Kreis)
ao_rd		Place of work: regional directorate (employment agency area)
wo_rd		Place of residence: regional directorate (employment agency area)
wo_kreis		Place of residence: district (Kreis)
ao_kreis		Place of work: district (Kreis)
ao_aa		Place of work: employment agency area
wo_aa		Place of residence: employment agency area
wo_aatyp06		Place of residence: employment agency region type 06
wo_ausl		Place of residence abroad

**Imprint****FDZ *Datenreport***

No. 03/2009 (EN)

**Publisher**

The Research Data Centre (FDZ)  
of the Federal Employment Service  
in the Institute for Employment Research  
Regensburger Str. 104  
D-90478 Nuremberg

**Editorial staff**

Stefan Bender, Dana Singula

**Technical production**

Dana Singula

**Copyright**

Reproduction – also in parts – only with permission of  
the FDZ

**Download**

[http://doku.iab.de/fdz/reporte/2009/DR\\_03-09-EN.pdf](http://doku.iab.de/fdz/reporte/2009/DR_03-09-EN.pdf)

**Internet**

<http://fdz.iab.de/>

**Corresponding author**

Jörg Heining, Tel.: +49 (0) 911/179-5392  
E-Mail: [Joerg.Heining@iab.de](mailto:Joerg.Heining@iab.de)