

Data Observer

Johanna Eberle and Alexandra Schmucker*

The Establishment History Panel – Redesign and Update 2016

DOI 10.1515/jbnst-2016-1001

1 Introduction

The Establishment History Panel (*Betriebs-Historik-Panel* – BHP) is one of the most comprehensive datasets on firms in Germany: it covers the population of establishments with at least one employee subject to social security. Only the data from the German Federal Statistical Office (Wagner 2015) and commercial data collections, such as the ORBIS database from Bureau van Dijk,¹ include similarly wide-ranging information on firms or establishments. The BHP is exceptional in that the data are not collected within a survey but come from an administrative source: the notifications to the German social security system. Every employer has to report every employment relationship subject to social security contributions at least once a year. The gathered data therefore contain information on all the employees concerned from the year 1975 onwards. These notifications also include a unique identifier for every employer, the establishment number, which is assigned by the German Federal Employment Agency (BA). Using this identifier, data obtained from individual employment notifications are aggregated to the establishment level to compile the BHP. The BHP includes the industry and location of the establishments as well as detailed information on the employment structure in terms of gender, age, qualification level, occupation, or average gross wages.

¹ See www.bvdinfo.com (accessed 15 November 2016).

*Corresponding author: **Alexandra Schmucker**, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Regensburger Str. 100, 90478 Nürnberg, Germany, E-mail: Alexandra.Schmucker@iab.de

Johanna Eberle, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Regensburger Str. 100, 90478 Nürnberg, Germany, E-mail: Johanna.Eberle2@iab.de

The first version of the BHP was released in 2007 by the Research Data Centre (FDZ) of the BA at the Institute for Employment Research (IAB) (Spengler 2008).² Since then the BHP has been updated with minor changes several times. For the current update, however, more substantial changes in the data preparation procedures and the catalogue of variables have been implemented for several reasons. First, adjustments in the social security notification system have led to modifications in the data reported on occupation, education and part-time work. Second, solutions developed by the IAB to improve the data quality, such as the imputation of earnings above the upper earnings limit and the validation of the information on education and vocational training, have been applied. Third, researchers' requirements and the increase in the individual data volume have been taken into consideration. As a result of this redesign, a new and enhanced BHP version for the years 1975 until 2014 has been created. The data preparation has been modified in certain respects in order to improve suitability for scientific research and may contribute to more comprehensive and precise research findings.

In this paper, we give an outline of the original data source, data preparation, and the contents of the final data set. In addition, we provide information on data access, working tools and the analysis potential of the BHP. A more detailed description of the BHP including frequency tables is provided by Schmucker et al. (2016).

2 Data source

2.1 Employee data

The BHP is based on individual data which are collected in the context of the employment notifications for the social security system. This procedure has been mandatory for all employees who are liable to contribute to the health, pension and unemployment insurances since 1973 in western Germany and since 1991 in eastern Germany. In 1999, the obligation to report was extended to marginal part-time workers. Employers have to submit notifications at the beginning and end of any employment relationship, once a year for continuous employment,

² Before that affiliated researchers of the IAB could use the Establishment File of the German Social Insurance Statistics (Fritsch/Brixy 2004) for their analyses, which was generated using the same data source. External researchers, however, were not entitled to use these data.

and whenever there are any changes relevant for social security.³ The notification contains information on the beginning and ending dates of the employment, gross wage, employment status, occupation, education and nationality. Characteristics of the establishment (location and industry classification) can be merged. These individual employment notifications are integrated into the Employee History of the IAB (*Beschäftigten-Historik* – BeH). The current BHP is based on a version of the BeH that comprises information on 1.5 billion employment episodes of 81 million individuals from 1975 until 2014.

2.2 Coverage

Although the BeH is a census, it does not cover total employment in Germany. Freelancers and civil servants are not subject to social security contributions and are therefore not included. Additionally, marginal part-time workers were not recorded in the data until 1998. With respect to the aggregate data set at establishment level, these restrictions mean that the data comprise only information on establishments with at least one employee subject to social security, or – since 1999 – at least one marginal part-time worker. As a result, a certain proportion of the working population and establishments are not observed in the BHP. This proportion is not evenly distributed across regions and industries and varies over time.⁴ As far as the industries are concerned, the manufacturing industry (excluding construction) and the private service sector (trade, logistics, warehousing, hospitality, IT and communications, finance, insurance, business services, and property) currently exhibit the

³ Within the employment notification procedure a certain time lag is unavoidable. Although changes in employment relationships have to be reported immediately and existing employment relationships have to be confirmed annually, some notifications actually arrive years later. It can thus be assumed that the number of employees and establishments is underrecorded slightly for the years 2012 and 2013 and that there are larger gaps for 2014. Analyses based on previous years suggest that about two per cent of the employees and three per cent of the establishments are still missing for 2014.

⁴ In order to determine what share of the total employment in Germany is included in the BHP data, the total number of employees per establishment as of 30 June 2013 is summed up for all establishments in the respective region or industry. These statistics are then compared to annual employment figures for 2013 obtained by the German Federal Statistical Office (Arbeitskreis 'Erwerbstätigenrechnung des Bundes und der Länder'), which also cover civil servants, the military and the self-employed (including family workers). The varying time reference is assumed to have little impact on the reported shares since seasonality is at a comparatively neutral stage in June, the reference date of the BHP.

highest coverage. For these industries, the BHP covers more than 85% of the total employment. For the construction industry, with its high share of self-employment, the coverage in the BHP amounts to only 78%. For the public service, defence and social security, which have a large share of public servants, the coverage is only 69%. In education, where the overwhelming majority of teachers are public servants, the share is 56%. The lowest shares are found for agriculture, forestry and fishing (51%) as well as for private households (41%) due to their high percentage of self-employment. The coverage may also vary significantly between regions. With respect to the sixteen federal states, coverage is lowest in Berlin (77%) and the other eastern German states (82%), and higher on average in western German states (85%) (see Figure 1). The varying coverage of employment needs to be given careful consideration when drawing statistical inferences.

2.3 Data quality

Administrative data have considerable advantages for researchers as they cover the total population and comprise precise and reliable information. Yet they are primarily collected for administrative purposes and research is only their secondary use. The suitability of the data for analysing research questions is not taken into consideration in the data collection process. Relevant information may therefore be missing: For example, exact information on gross wages is only provided up to the upper earnings limit for statutory pension insurance contributions. All earnings above this limit are censored at that limit. In addition, changes in the data collection method and the recorded information may result in inconsistent time series. Several approaches to deal with biases and inconsistencies have been implemented in the BHP (see chapter 3.1).

For the current BHP version we had to deal with a substantial change in the information gathered: the occupation code is an array where employers give information on the employee's occupation, school and vocational training, and other job details. In 2011, the former recording scheme was superseded by a new occupation code in the social security notification procedure (see Bertat et al. 2013). Table 1 compares the contents of the new and the old occupation codes. In addition to the changes in the categories for education and occupation, it is to be noted that the new occupation code no longer records any information on occupational status (see Table 1 for more information on this variable), but now contains information on agency work and fixed-term contracts.



Figure 1: BHP coverage of total employment in 2013 (%), German federal states.

Source: Authors' calculations based on Establishment History Panel 7514 and annual employment figures by Arbeitskreis Erwerbstätigenrechnung des Bundes und der Länder; Geographic allocation © GeoBasis-DE/BKG 2015.

Moreover, there was a transitional period of six months during which the verification programs of the notification procedure permitted the omission of information on the occupation codes. About 20% of the notifications for this period therefore do not include valid information on the occupation code. It can be seen that the number of notifications with missing information is not distributed randomly, but differs systematically across regions and sectors of the economy (for more details see Schmucker et al. 2016). For some of these changes, missing-data solutions were developed and applied during the creation of the BHP (see chapter 3.1).

Table 1: Contents of the new and the old occupation codes.

Contents	Values of old occupation code	Values of new occupation code
Occupation performed	Classification of Occupations 1988 (KldB 1988)	Classification of Occupations 2010 (KldB 2010)
Occupational status	Trainees/apprentices; unskilled employees; skilled workers; master craftsmen, foremen; home workers and freelance home workers	Not recorded
Part-time	Full-time; part-time (less than 18 hours); part-time (18 hours and more)	Full-time; part-time
Education and vocational training	<ul style="list-style-type: none"> – Lower/intermediate secondary school leaving certificate without completed vocational training – Lower/intermediate secondary school leaving certificate with completed vocational training – Upper secondary school leaving certificate without vocational training – Upper secondary school leaving certificate with vocational training – Degree from a university of applied sciences (<i>Fachhochschule</i>)– University degree – Education and training unknown, no statement possible 	Education: <ul style="list-style-type: none"> – None – Lower secondary school leaving certificate – Intermediate school leaving certificate or equivalent school education – Upper secondary school leaving certificate (general or subject-specific aptitude for higher education) Vocational training: <ul style="list-style-type: none"> – No vocational qualification – Qualification from a recognised vocational training programme – Master craftsman, technician, or equivalent qualification from a technical college – Bachelor degree – Diploma, Magister, Master's degree, state examination – Doctorate
Agency worker	Not recorded	Yes; no
Fixed-term contract	Not recorded	Permanent; fixed-term

3 Data preparation

3.1 Pre-processing of the employee data

The BHP is compiled by aggregating the employment notifications of the BeH at establishment level using the establishment ID. Before this aggregation, the data

on individuals are subjected to numerous validation procedures. The most important improvements affect the variables on income, education and part-time work, which are described in the following.

Information on earnings in the social security notifications is very precise and valid in general. However, earnings are only reported up to the upper earnings limit for statutory pension insurance contributions. For this reason, approx. 10 % of the data on full-time employees' earnings is censored. This leads to biased results due to aggregation at the establishment level. In order to remedy this bias, the information on earnings was imputed following the procedure developed by Card et al. (2015) (see also Schmucker et al. 2016: 125).

The number of employment notifications with missing information on education and vocational training qualifications had grown substantially over time. The switch to the Occupation Code 2010 in the notification procedure then caused the rate of missing values to rise as high as 50 % in 2011. Furthermore, from 2011 onwards the employers no longer report qualifications in a combined variable, but split into school education and vocational education and training (see Table 1). Hence no time-consistent information is available for the entire period. In order to provide researchers with time-consistent information on school and vocational education, every combination of values from the new occupation code is assigned to the most suitable value of the combined education and vocational training variable according to the old occupation code. This has no effect on missing values, however. In addition, the quality of the education and training data is improved by means of an imputation procedure using a deterministic replacement rule that was suggested by Fitzenberger et al. (2005, 2006) and enhanced by Kruppe et al. (2014). The result of this procedure is that there are now hardly any missing values, especially for employees who are not in marginal part-time employment (see Schmucker et al. 2016: 115ff).

As mentioned above, the possibility to leave out the information on the occupation code for six months after the introduction of the new occupation code resulted in many missing values in the data. The information regarding working hours (full-time or part-time) was also affected. This problem was crucial for the creation of the BHP as the full-time employees are often used as the basis for generating employment groups in the BHP. In order to remedy this problem a logit model was developed at the IAB which can be used to impute the missing information (see Ludsteck/Thomsen 2016).

The Classification of Economic Activities has changed several times during the observation period of the BHP. Inconsistencies in the classification scheme make it difficult to conduct longitudinal analyses. To address this issue, the IAB developed a method to construct time-consistent industry codes. A detailed

description can be found in Eberle et al. (2011). These variables are available in the BHP in addition to the original classifications.

A similar problem arises due to local government reorganisations throughout the entire period. The regional information on districts (*Kreise*) was therefore corrected according to the territorial allocation as of 31 December 2014. A more detailed description of all validation procedures can be found in Schmucker et al. (2016: 22 ff.).

3.2 Compiling the establishment data

After the data on individuals have been pre-processed, the BHP is compiled by aggregating the employment information at establishment level by using a unique establishment identifier. Each establishment is allocated a unique establishment number according to the following principle: branch offices of one company which belong to the same economic class and are located in the same municipality are given one joint establishment number (see Bundesagentur für Arbeit 2007). As the identification of establishments in the BHP is based on these establishment numbers, the definition of an ‘establishment’ is specific and may differ from other concepts. It is not possible to distinguish between branch offices with a joint establishment number in the data. It should also be noted that there is no information available on whether establishments belong to the same company.

The final BHP contains information on the establishments' employee structure and details about the wage structure of the full-time employees in the establishment. All of the data are stock data as of the reference date of 30 June of any given year.

In addition to details about the stock, the BHP also includes information on worker inflows and outflows. Inflows are defined as the number of employees who were working in the establishment on the respective reference date but were not working there on the reference date of the previous year. A corresponding definition applies to outflows. Similar to the stock values, inflows and outflows can be distinguished according to various characteristics such as age groups.

Finally, the BHP has been enriched with information on establishment start-ups and closures. A new establishment ID, which defines an establishment in the BHP, does not necessarily imply that a whole new establishment has been set up. Also, the disappearance of an establishment ID does not necessarily mean that the establishment has closed down. The classification provided helps to distinguish between genuine start-ups and closures and changes in

establishment numbers, takeovers or spin-offs. Establishment entries and exits are classified by analysing worker flows. The decisive factor here is the proportion of workers who were employed in one and the same establishment in the previous or in the subsequent year. A detailed description of the classification can be found in Hethey and Schmieder (2010).

4 Data availability

4.1 The BHP sample

The BHP is made available to the scientific community via the FDZ. Access is granted to a 50 % simple random sample of all establishments included in the BHP from 1975 to 2014. Due to the random selection, the sample shows no systematic bias. The BHP data are provided as Stata data sets and contain both German and English labels.

Figure 2 shows the number of establishments per year in the BHP sample over the period 1975–2014. There are two remarkable rises in 1992 and 1999, which are highlighted by two vertical lines. The first is caused by the inclusion

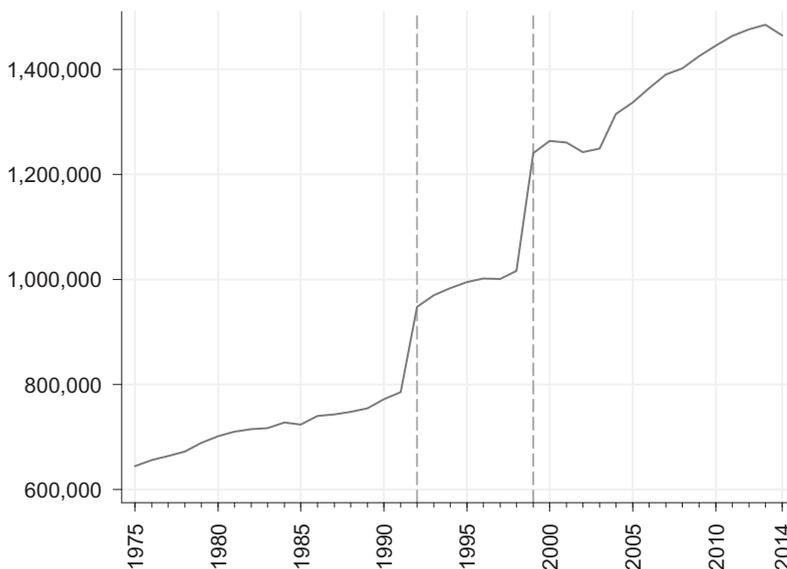


Figure 2: Number of establishments per year in the BHP 50 % sample.

of establishments in eastern Germany, the second is due to the integration of marginal part-time workers in the notification procedure.

4.2 Availability of BHP information in FDZ data products

Selected information drawn from the BHP can also be merged with other data products provided by the FDZ. For example, BHP modules are available for the IAB Establishment Panel (Fischer et al. 2009), a comprehensive survey of establishments in Germany conducted by the IAB. BHP data on employers is also available for all individuals covered in the Sample of Integrated Labour Market Biographies (SIAB) (Antoni et al. 2016), a data set comprising information on employment, unemployment benefit receipt, registered job search, and participation in training schemes. The SIAB data include an establishment file holding basic information on these establishments (total number of employees, full-time and marginal part-time workers, mean wage, industry, and region). In addition to basic establishment characteristics, several modules and blocks of variables can be provided upon request. Selected BHP modules are also available for many other FDZ data products.⁵ By adding administrative information on establishments, FDZ aims to meet a growing demand for combined research data sets providing data on both employees and employers.

4.3 Data access

Access to the BHP is only possible via on-site use at the FDZ and remote data processing. For on-site use, the FDZ provides several workplaces for visiting researchers within a secure computing environment at different locations in Germany and abroad (Bender/Heining 2011).⁶ For remote data processing, researchers submit scripts to the FDZ that execute commands for data preparation and analysis. The scripts are processed on secure servers within the FDZ and the results are transferred to the researchers after a disclosure review. Remote data processing at the FDZ is conducted via the Job Submission Application (JoSuA) developed by the Institute for the Study of Labour (IZA),

⁵ For a list of available modules please see http://doku.iab.de/fdz/access/BHP_Variablen_EN.pdf (accessed 24 November 2016).

⁶ For an up-to-date list of external FDZ locations, see http://fdz.iab.de/en/FDZ_Data_Access/FDZ_On-Site_Use/Standorte.aspx (accessed 23 November 2016).

which allows users to submit jobs and access their results via a custom-built web interface (see Eberle et al. forthcoming).

Researchers intending to use the BHP are required to apply for data access via on-site use and remote data processing at the FDZ. Access can be granted to non-commercial research institutions conducting research projects in the fields of employment research.

After the request for data access has been accepted, the FDZ and the research institution conclude a data use agreement. The data can only be used for a specific research project and within the time period stated in the contract. Students intending to use the BHP for their theses are also welcome to apply for the data. In this case, the data use agreement is concluded with the department supervising the thesis.

Since direct access to the data is only possible during on-site use, the FDZ provides artificial test data sets for download on its website, which can be used by researchers to develop scripts for data preparation and analysis and to test their functionality.⁷ The test data were designed in such a way as to resemble the structure of the real data as closely as possible, but do not allow any valid analyses. Values are switched randomly between observations within certain subgroups. Test data are also smaller in size since they are based on a subsample.

5 Analysis potential

The BHP facilitates a wide range of studies in the field of labour market research. In the past it has been the basis for analyses on start-ups and closures as well as on the survival of newly founded firms (Brixy 2014; Fackler 2014). Additionally, projects on the development of specific regions (Braakmann/Vogel 2011) or industries (Bossler/Oberfichtner 2016) have used the BHP. Finally, analyses on employment adjustment over the business cycle (Bossler/Upward 2016) or on (cultural) diversity (Brunow/Nijkamp 2016) have also been possible.

6 Conclusion

As demonstrated in this article, the BHP is a rich administrative data set with longitudinal information on establishments in Germany from 1975 to 2014. It was

⁷ See http://fdz.iab.de/en/FDZ_Establishment_Data/Establishment_History_Panel/Establishment_History_Panel_Working_Tools.aspx (accessed 24 November 2016).

designed to make a multitude of research projects in the fields of employment research possible. More detailed information on the BHP can be found in the corresponding data report (Schmucker et al. 2016). A list of relevant articles concerning the BHP, such as data reports, method studies and research papers using the BHP can be found by navigating through the FDZ literature database.⁸ For the latest news on the BHP, including updated versions, please refer to the FDZ website fdz.iab.de.

Acknowledgements: The authors would like to thank Rebecca Hussen for her assistance in assembling statistics for this article and fellow staff members at the IAB who were involved in the compilation of the BHP.

References

- Antoni, M., A. Ganzer, P. vom Berge (2016), Sample of Integrated Labour Market Biographies (SIAB) 1975–2014. FDZ-Datenreport 04/2016 (en).
- Bender, S., J. Heining (2011), The Research-Data-Centre in Research-Data-Centre Approach: A First Step Towards Decentralised International Data Sharing. FDZ-Methodenreport 07/2011 (en).
- Bertat, T., A. Dundler, C. Grimm, J. Kiewitt, C. Schomaker, H. Schridde, C. Zemann (2013), Neue Erhebungsinhalte „Arbeitszeit“, „ausgeübte Tätigkeit“, sowie „Schul- und Berufsabschluss“ in der Beschäftigungsstatistik. Methodenbericht der Statistik der Bundesagentur für Arbeit. Available at: <http://statistik.arbeitsagentur.de/Statistischer-Content/Grundlagen/Methodenberichte/Beschaeftigungsstatistik/Generische-Publikationen/Methodenbericht-Neue-Erhebungsinhalte-Arbeitszeit-ausgeuebte-Taetigkeit-sowie-Schul-und-Berufsabschluss-in-der-Beschaeftigungsstatistik.pdf> (Accessed: 15.11.2016, only available in German).
- Bossler, M., M. Oberfichtner (2016), The Employment Effect of Deregulating Shopping Hours. Evidence from German Food Retailing. *Economic Inquiry*, online first.
- Bossler, M., R. Upward (2016), Employee turnover and the expansion and contraction of employers. PP. 305–346 in: G. Saridakis, C. L. Cooper (Eds.), *Research Handbook on Employee Turnover*, Cheltenham, New Horizons in Management.
- Braakmann, N., A. Vogel (2011), How Does Economic Integration Influence Employment and Wages in Border Regions? The Case of the EU Enlargement 2004 and Germany’s Eastern Border. *Review of World Economics* 147(2): 303–323.
- Brixy, U. (2014), The Significance of Entry and Exit for Regional Productivity Growth. *Regional Studies* 48(6): 1051–1070.

⁸ See http://fdz.iab.de/en/FDZ_Publications/FDZ_Literature_Database.aspx (accessed 24 November 2016).

- Brunow, S., P. Nijkamp (2016), The Impact of a Culturally Diverse Workforce on Firms' Revenues and Productivity. An Empirical Investigation on Germany. *International Regional Science Review*, online first.
- Bundesagentur für Arbeit (Ed.) (2007), *Handbuch für die Betriebsnummernvergabe und -pflege im Rahmen des Meldeverfahren zur Sozialversicherung*. Edition December 2007 (only available in German).
- Card, D., J. Heining, P. Kline (2015), *CHK Effects*. FDZ-Methodenreport 06/2015 (en).
- Eberle, J., P. Jacobebbinghaus, J. Ludsteck, J. Witter (2011), *Generation of Time-Consistent Industry Codes in the Face of Classification Changes: Simple Heuristic Based on the Establishment History Panel (BHP)*. FDZ-Methodenreport 05/2011 (en).
- Eberle, J., D. Müller, J. Heining (forthcoming), *A modern job submission application to access IABs confidential administrative and survey research data*. In: *Statistics Canada (Ed.), Statistics Canada International Symposium 2016: Growth in Statistical Information: Challenges and Benefits*.
- Fackler, D. (2014), *Establishment Survival in East and West Germany. A Comparative Analysis*. *Schmollers Jahrbuch* 134(2): 183–208.
- Fischer, G., F. Janik, D. Müller, A. Schmucker (2009), *The IAB Establishment Panel * Things Users Should Know*. *Schmollers Jahrbuch* 129(1): 133–148.
- Fitzenberger, B., A. Osikominu, R. Völter (2006), *Imputation Rules to Improve the Education Variable in the IAB Employment Subsample*. *Schmollers Jahrbuch* 126(3): 405–436.
- Fitzenberger, B., A. Osikominu, R. Völter (2005), *Imputation Rules to Improve the Education Variable in the IAB Employment Subsample*. FDZ-Methodenreport 03/2005 (en).
- Fritsch, M., Brixy, U. (2004), *The Establishment File of the German Social Insurance Statistics*. *Schmollers Jahrbuch* 124(1): 183–190.
- Hethey, T., J.F. Schmieder (2010), *Using Worker Flows in the Analysis of Establishment Turnover – Evidence from German Administrative Data*. FDZ-Methodenreport 06/2010 (en).
- Kruppe, T., B. Matthes, S. Unger (2014), *Effectiveness of Data Correction Rules in Process-Produced Data: The Case of Educational Attainment*. IAB-Discussion Paper 15/2014.
- Ludsteck, J., U. Thomsen (2016), *Imputation of the Working Time Information for the Employment Register Data*. FDZ-Methodenreport 01/2016 (en).
- Schmucker, A., S. Seth, J. Ludsteck, J. Eberle, A. Ganzer (2016), *Establishment History Panel 1975–2014*. FDZ-Datenreport 03/2016 (en).
- Spengler, A. (2008), *The Establishment History Panel*. *Schmollers Jahrbuch* 128(3): 501–509.
- Wagner, J. (2015), *25 Jahre Nutzung vertraulicher Firmenpaneldaten der amtlichen Statistik für wissenschaftliche Forschung: Produkte, Projekte, Probleme, Perspektiven*. *Wirtschafts- und Sozialstatistisches Archiv* 9(2): 83–106.