

Daniela Hochfellner, Tatjana Mika, Dana Müller, Michael Stegmann

**BASiD: “Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland” (Biographical data of selected social insurance agencies in Germany) - Constructing a longitudinal-biographical data set, based on the data of the German Pension Fund the Federal Employment Agency and the Institute for Employment Research**

**Aims**

For the first time this project compiles individual-level data of two dataset holding institutions to one biographical dataset. Large parts of the individual data of the Federal Employment Agency (BA) and the Institute for Employment Research (IAB) as well as the individual data of the German Pension Insurance (RV) come out of the same data generating process and are built on the equal legal basis namely the Code of Social Law. The ambition of this project is to compile a common dataset which contains data of the RV and the BA respectively the IAB. For Researchers this data set will be offered either as Scientific Use File (SUF) or in a weakly anonymous version for one-site-use.

**Generating the dataset/ workflow**

The Completed Insurance Biographies (VSKT)<sup>1</sup> of the RV can be regarded as the basic file for this project. Accordingly all persons who are in this sample will be identified in the data of the IAB. Data sources for this identification are the Employee and Benefit Recipient History (BLH)<sup>2</sup>, the Integrated Employment Biographies (IEB) and the Establishment History Panel (BHP). By matching these data, information from the first contribution until the point of the latest available data will be obtained, for persons between 15 and 67 years of age who are liable to social security.

In the primary project phase the variables, existing in both data sets, are merged and tested. There will be difficulties with merging the data, resulting on the one hand out of the fact that the information is not existing in the same form in both institutions, on the other hand out of the fact that the existence of parallel states could be possible. For example the official unemployment statistic of the BA differs from the data of the RV, because the two institutions collect this data in a different way. To handle these problems suitably heuristics have to be developed. Therefore the data review and -preparation is related to a high expenditure of time.

After data cleansing and summarizing the parameters/labels due to reasons of data security, the characteristics beyond the common variables of both datasets, will be merged. In the following step a proposal for the SUF will be presented to the data security. The developed dataset should eventually be offered for test use to find possible improvement opportunities for the final version of the SUF. Research questions on regional backgrounds are getting more and more important. Because of this a feasibility-study for constructing a SUF for regional analyses will be done additionally. In order that foreign researchers should be able to use the SUF, the documentation within the project is kept in German and English.

---

<sup>1</sup> The VSKT is a permanent special survey of the German Pension Insurance. It includes all the insured persons of the German Pension Insurance and their state of retirement pension expectancy in Germany. Further it provides information about all (stored) pension-related facts, that means information about legal times of pension and contributions are stored after an account declaration in the insurance account.

<sup>2</sup> The BLH includes a complete history of the employees subject to social security and individuals with benefit payments since 1975. The IEB contains useful information about the employees subject to social security and individuals with benefit payments as well as references to times of job-seeking and measurement participation. The BHP includes all establishments in Germany with at least one employee liable to social security at the 30<sup>th</sup> of June each year. Since 1999 establishments with at least one marginal part-time employee are additionally enclosed.

**Characteristics of the dataset:**

The new dataset will differ in certain characteristics from the previous existing datasets. It contains on the one hand more characteristics, which allow to deal with research questions that could not be answered for Germany or just less precise in the present, on the other hand the dataset displays the complete biography of the individuals also before 1975 and is not left-censored for any individual.<sup>3</sup>

Besides the data-technical and demographical/regional characteristics there are some other variables of the RV, like pension times, consideration times, allowance times, payment dates and birth dates for children of the individuals in the SUF. The contained data of the BA respectively the IAB are establishment-information, like the establishment-size, classification of industries and information gained from the social security notification process. This information implies apprenticeship-, employment-, unemployment-, job-seeking-, training- and payment-details.

For example analyses with regard to:

- birthrates and employment histories of women
- resettles
- diseases and care over the employment cycle
- the influence of military/ civil service on the employment histories
- life-income and earning points for the pension
- influence of start-up-conditions on the career
- changes on law and initiation points and
- obtaining single regions/ federal states

can be arranged.

Both data research centers like to work on own research projects with the new developed dataset, like the feasibility-study for the SUF for regional analyses. These research projects are necessary to identify advantages and disadvantages of the dataset and to include detected improvements.

**Data supply and access**

Researchers in Germany and in foreign countries can obtain the dataset from both research data centers without being charged. The data are mainly offered as SUF, because it is still the preferred way for Researchers of accessing a dataset and it can also be easily transmitted abroad. Furthermore to allow more detailed analyses (regions or trainings activities types) a weakly anonymous version will be provided for one-site-use.

Information:

Information is available on the homepage of the FDZ of the BA/IAB on the link below:

[http://fdz.iab.de/de/FDZ\\_Projects.aspx](http://fdz.iab.de/de/FDZ_Projects.aspx)

---

<sup>3</sup> That is already the circumstance for the datasets offered by the RV, but it is not valid for the data of the Research Data Centre of the BA in the IAB, which are introduced for the first time in 1975.