Schmollers Jahrbuch 129 (2009), 501 – 513
Duncker & Humblot, Berlin

European Data Watch

This section offers descriptions as well as discussions of data sources that are of interest to social scientists engaged in empirical research or teaching courses that include empirical investigations performed by students. The purpose is to describe the information in the data source, to give examples of questions tackled with the data and to tell how to access the data for research and teaching. We focus on data from German speaking countries that allow international comparative research. While most of the data are at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) are included as well. Suggestions for data sources to be described in future columns (or comments on past columns) should be send to: Joachim Wagner, Leuphana University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to ⟨wagner@leuphana.de⟩. Past "European Data Watch" articles can be downloaded free of charge from the homepage of the German Council for Social and Economic Data (RatSWD) at: http://www. ratswd.de.

Official Firm Data for Germany

By Anja Malchin and Ramona Voshage

Motivation

The research data centre (RDC) of the statistical offices of the Länder – founded in 2001 – has through the last years become a well-established infrastructural facility along with the other research data centres in Germany. Safe scientific workstations in all of the statistical offices of the Länder and in independent scientific institutions as well as the possibility to use microdata from a wide variety of official statistics are part of the options of this regional data infrastructure. These advantages resulted in a high request of the offered services by the scientific community (Zühlke et al., 2007; Kaiser / Wagner, 2007). After a first project phase up to the end of June 2007 the RDC of the statistical offices of the Länder is now funded by the Federal Ministry for Education and Research (BMBF) for a second project phase, which integrates new

challenges for the regional offices of the RDC. One important task is the further development of the provided data sets and the range of services offered according to the expectations of the scientific community. The first chapter of this paper will inform about the origins of the sub-project "Official Firm Data for Germany (AFiD)"[1], followed by a description of the project content. The third chapter will cover the data sets, which will be created during the course of the project. This paper will be completed by a discussion of the future prospects of the RDC.

1. Project description

The most important task of the RDC is to supply the scientific community with official microdata within the scope of the legal requirements. In order to do so, the staff of the RDC has prepared microdata sets for various statistics and made them available for scientific purposes by different access paths (see www.forschungsdatenzentrum.de). Part of this newly created data pool are microdata from units and enterprises[2]. The data is being collected by the statistical offices at regular intervals in single branches of economic activities, most of the statistics are compulsory. The created data sets are available only for individual branches and as cross-sectional data. Compared to other sectors the branch of manufacturing and mining and quarrying is well covered by official statistics. In this sector several surveys are conducted in establishments as well as in enterprises. Besides this also data from the agriculture or the service sector and recently also data from the branch of electricity, gas and water supply are frequently asked for by the scientific community. Another well used data source is the survey of salary and wage structure resp. structure of earnings statistics. For this survey establishments from various branches have to respond.

The RDC has made firm data available for scientific analyses for quite some time by different access paths (Malchin / Pohl, 2007). However, there were also a few requests for combined firm data (Görzig et al., 2005; Görzig / Gornig, 2005). For these projects the staff of the RDC merged the data according to the needs of the specific scientific project to find that this process of merging was rather time-consuming and needed a lot of labour. But because the information potential of the firm data could be considerably increased by merging and because there was obviously an interest in these merged data sets by the scientific community, the basic idea of the project AFiD was born.

[1] The acronym stands for the german *"A*mtliche *Fi*rmendaten für *D*eutschland".

[2] In official statistics a unit or establishment is defined as the smallest local unit, the enterprise being the smallest legal entity. In this paper the official statistics concerning establishments as well as enterprises are referred to as firm data.

2. Project content

For prompt and flexible merging of project-defined firm data a sufficient data management concept has to be formed. The ambition of AFiD is the development of an extensive treasure of data about units and enterprises in Germany, allowing mergings of economic and environmental microdata for longitudinal analyses and integration of these data sets using the business register (URS[3]). Save for a few exceptions the URS forms the sampling frame for most of the economic statistics (Brandt et al., 2008). It exists since 1998 and covers all economic entities contributing to the gross domestic product, having their place of business in Germany and the economic focus in the according business sector (Statistisches Bundesamt, 2006). From the year 2001 on it's possible to merge firm data coming to valid results (Sturm / Tümmler, 2006).

The legal base for merging of economic and environmental microdata has been formed with the amendment of the federal statistics law (BStatG[4]) in the year 2005. According to the inaugurated § 13a BStatG unit- and enterprise data from different sources of official statistics are now allowed to be integrated as far as it's aimed to gain statistical information without conducting additional statistical surveys.

When merging different data sources and integrating microdata using the URS various peculiarities have to be taken in consideration. On one hand the technical feasibility of data combination has to be verified, i.e. there have to be identifiers in both data sets, which allow merging and which are permitted to use for this purpose by law. Also the level of investigation has to be kept in the dataset: in some surveys single units are questioned, in some surveys enterprises. Generally it's possible to aggregate from unit level to enterprise level. Note that the enterprise will only be fully represented, when the original data covers information for all local units. As soon as only one local unit of an enterprise has an economic focus in another sector of economic activities or not all local units of the enterprise are part of the survey, important information to the enterprise is missing in the data set. On the other hand the varying survey design has to be considered. While in some surveys all units or enterprises up to a certain cut-off are questioned, other surveys can be rotating sample surveys.

Because of the complexity of this process the merging will be done stepwise. In the first step longitudinal datasets for single and combined branches of economic activities are developed. By this time some of these panel datasets can already be analysed at the safe scientific workstations and per remote data processing (on-site-use). The integration of selected data sets to the URS

[3] URS meaning Unternehmensregister, see Verordnung Nr. 2186 / 93 (EU-Register-Verordnung).

[4] BStatG meaning Bundesstatistikgesetz.

will follow in a second step. The current project status and detailed information to all products can be found at the homepage of the RDC www.forschungsdatenzentrum.de / afid.asp.

3. Products

The AFiD-products can be grouped into paneldata and modules. The paneldata contains cross-sectional surveys, which are pooled to longitudinal data material as well as merged data from similar groups of respondents. Modules either contain data material from multi-annual statistics, partly covering information to several branches of economic activities or are suitable for specific research questions, e.g. environmental analysis.

Similar to an add-on system the separate or joint data material can be applied for by the scientific community and will be made available for analysis in a flexible and prompt manner. In addition, this separation of the products into paneldata and modules helps to avoid redundancies when merging the created datasets to the business register. Some of the products are described below.

3.1 Paneldata

The created paneldata differ by the observed unit level: there are paneldata on the level of the local production unit and enterprises level data. This phenomenon depends on the particular statistics. Merging is always done on the level which is analogue to the group of respondents (on unit level for unit-related surveys, on enterprise level for enterprises-related surveys). Table 1 in the appendix provides a compact overview of the AFiD-paneldata. The available paneldata are sub-divided according to the following branches of economic activities.

3.1.1 Agriculture

By creating an *"AFiD-Panel Agriculture"* the information potential of the currently available agricultural statistics has been improved considerably. The data combination followed a stepwise procedure. In a first step the microdata of the census of agriculture 1999 and the census of agricultural structure 2003 and 2007 were merged to a longitudinal data set. The representative statistics of horticulture, land use and live stock (cross-sectional data) are to be included in the panel during the second step. Since every agricultural holding in Germany is given a unique ID-number, both steps become realisable using the agricultural register.

Agriculture in Germany faces great challenges nowadays by liberalization of the EU-markets, rising requirements for product quality and product safety

and the opening of new markets like bioenergy. The new AFiD-Panel Agriculture gives the possibility to observe developments like this over the time. Especially for research in the history of Germanys agricultural structure a valid longitudinal data base will be a great asset. If required, other statistics (e.g. the vegetable production survey) can be merged to the existing paneldata set.

3.1.2 Manufacturing and Mining and Quarrying

Production units and enterprises having their focus in the branch of manufacturing[5] are respondents to a bundle of alike surveys, having to provide information about their turnover, employment, cost structure, investments or products. For this branch of economic activities two sets of paneldata will be created, according to the observed unit level of the respondents: "AFiD-Panel Industrial Units" und "AFiD-Panel Industrial Enterprises".

The *"AFiD-Panel Industrial Units"* expands the establishment panel (Konold, 2007), adding information from the quarterly production survey. The data set will contain annual results from the monthly report and the production survey, annual data from the survey of investments and the longitudinal data of the survey of small units since the year 1995. These surveys are census surveys with a cut-off limit and cover a similar group of respondents. Normally industrial units with at least 20 employees are polled. An exception is the survey of small units, containing information of establishments having not more than 19 employees.[6]

The AFiD-Panel Industrial Units covers information about employment, wages and salaries, hours worked and turnover (distinguished into domestic and non-domestic), sale production value and investments in acquired and self-provided fixed assets (e.g. real estate), all on the level of the production unit.

This panel accumulates to a global longitudinal data set for units in manufacturing, being still expandable using the modules also developed in the AFiD-Project. For a description of the research potential of these microdata see Wagner (2007).

For enterprises with their economic focus in manufacturing the *"AFiD-Panel Industrial Enterprises"* covers the essential information collected by official

[5] The existing statistics cover the branch of manufacturing as well as mining and quarrying. In order to simplify matters in the following only the terms manufacturing or industry will be used.

[6] The survey of small units has been stopped in the year 2002. From the reporting year 2003 on, the group of respondents decreases to production units being part of industrial enterprises with at least 20 employees (Malchin / Pohl 2007). Further information to the other surveys are to be found in the quality reports of the Federal Statistical Office (Statistisches Bundesamt, 2005 a, c, d).

statistics. Ultimate aim is the combination of data in a form that allows cross-sectional as well as longitudinal analyses. The data set contains microdata from the annual survey for (multi-unit) enterprises and from the investment survey for enterprises combined with the cost structure survey (KSE[7]) in manufacturing for the reporting years 1995 to 2006 (Statistisches Bundesamt 2005b).

Apart from the KSE these surveys are census surveys with a cut-off limit. The KSE is a sample survey[8]. To disburden the enterprises obliged to provide information, new sampling is normally done every four years. The panel contains industrial enterprises usually having at least 20 employees (Statistisches Bundesamt 2005d).

The AFiD-Panel Industrial Enterprises covers basic information about all enterprises in manufacturing, in particular about turnover, persons in employment and total wages and salaries as well as details to in- and outflow of fixed assets (e.g. real estate or machines and equipment), combined with data from the KSE, containing complementing information about production results, value of production factors and added value.

The AFiD-Panel Industrial Enterprises allows analyses of the developments in manufacturing on the level of the smallest legal entity. Due to the implementation of the KSE-microdata this data set is suitable for various structural studies (Fritsch et al., 2004).

3.1.3 Electricity, gas and water supply

Similar to the branch of manufacturing also in the field of energy different surveys are being conducted on the level of both establishments and enterprises. These existing microdata have been merged to a longitudinal establishment data set as well as to an enterprise panel.

In the *"AFiD-Panel Energy Units"* annual results of the monthly report and microdata from the investment survey for establishments in energy- and water supply are included. These two surveys have the same group of respondents, on one hand due to the structure of the observed line of business and on the other hand due to the defined cut-off threshold to units with 10 or more employees (including establishments owned by public authority).

The panel covers the years 2003 to 2006 and contains employee-related information like the number of employees, hours worked and wages and salaries

7 KSE meaning Kostenstrukturerhebung.

8 The enterprises are selected in a stratified random sample, using the branch of economic activities and employee size class as stratification characteristics. Sampling frame is the german business register, sample size is about 18.000 enterprises. Please note that in the KSE enterprises of some strata (from very important branches and employee size classes) are completely polled.

as well as business-related information like investments in fixed assets or real estate and sales revenue. Concerning investments in environment protection the units are questioned, if in the details given also equipment for prevention, disposal or reduction of harmful environmental intervention is included. Using the information of the investment survey it's also possible to allocate the establishments to their main line of business (supply with electricity, heating, gas or water).

For the *"AFiD-Panel Energy Enterprises"* microdata from the cost structure survey and the investment survey were merged to a longitudinal data set on the level of the enterprises in the field of energy- and water supply (starting with the year 2003).

Again these are census surveys with a cut-off limit of 10 or more employees, so that almost all enterprises in this branch including the enterprises owned by public authority are obligued to submit information.

The cost structure survey provides information about employees, turnover, material and stocks, procurement and consumption, costs by type of costs, taxes and subsidies as well as information about in-house research and developments and supply with water. More enterprise-related data derives from the investment survey: legal form, tax group status and line of business (electricity generation from thermal power or other sources, heat / gas / water supply, sewage disposal). Furthermore the survey covers gross input, values of fixed assets and real estate and the according revenues. Similar to the units the enterprises also have to state if any equipment for prevention, disposal or reduction of harmful environmental intervention is included in the given details.

3.1.4 Services

During the last years a lot of enterprises were founded in the service sector and this branch gains more and more importance for the whole economy. Using the relatively new structural survey in the services sector enterprises can now be analysed e.g. in terms of growth over a time period which spans several years.

The *"AFiD-Panel Services"* covers information about the number and characteristics of persons employed, wages and salaries, turnover, investments, taxes and subsidies in enterprises either below or above an annual turnover of 250.000 Euro (Vogel, 2009). These variables are derived from the structure survey in the services sector. This sample survey covers about 15% of all units and enterprises registered in the URS and is conducted yearly.[9] The surveys were combined to a longitudinal data set covering the reporting years 2003 to 2006.

[9] More details to this survey are to be found in the latest quality report of the Federal Statistical Office (Statistisches Bundesamt, 2008).

3.1.5 Business Register

For the new *"AFiD-Panel Business Register"* the cross sections of the years 2004 to 2007 were combined. By creating a longitudinal data set for the business register the options for scientific research will be considerably extended. In the URS included are all active enterprises having taxable trade accounts turnover and / or employees subject to social insurance contributions in the reporting years 2002 to 2005. Apart from agriculture, hunting and forestry, fishing and public administration units and enterprises from all branches are nearly completely recorded.

The longitudinal data set covers amongst others information about the main economic activity of the firms, the number of employees, turnover and corporate affiliation, therefore allowing widespread analyses of economic structures or consolidation processes over time.

3.2 Modules

For some of the created paneldata the information potential can be enlarged even further by adding the already available AFiD-Modules (see table 1 in the appendix). So the AFiD-Panel "Industrial Units" can be merged with the modules "Earnings" and "Use of Energy" or with the environmental modules "Water Supply and Wastewater Disposal", "Environmental Protection Expenditures" or "Environmental Protection Commodities". So prospectivly, official firm data not only allows analyses of demographic developments in connection with cost items, but also in connection with characteristics from the environmental or energy sector.

The created AFiD-Modules are specified as follows.

The *"AFiD-Module Earnings"* includes the combined structure of earnings surveys[10] of the years 2001 and 2006. The structure of earnings survey is a linked employer-employee-data set and covers information both about the units (amongst others about ownership ratio of public authority) and the employed persons (especially the kind of work, the earnings and if there is a collective agreement giving the frame for the salary). By combining the AFiD-Module Earnings with the AFiD-Panel Industrial Units especially the informations on the establishment-level can be enlarged.

Another extension for the AFiD-Panel Industrial Units is the *"AFiD-Module Use of Energy"*. On establishment-level information about the volume of purchased and delivered electricity (from electricity providers, from other companies, from foreign companies) as well as about power generation (by hydroelectricity, thermal power or other sources) are available. Details to purchase

[10] More information are to be found in Hafner / Lenz, 2007.

and usage as well as delivery and stock of fuel are partially collected by energy source.

The three environmental AFiD-Modules are each based on an environmental survey. The group of respondents is made up mostly by establishments in manufacturing. Detailed information to the individual modules can be found at the internet presence of the research data centre of the statistical offices of the Länder: http://www.forschungsdatenzentrum.de/afid.asp.

Using the AFiD-Module *"Water Supply and Wastewater Disposal"* researchers now have the opportunity to analyse features relevant to the environment (like water usage or recycling activities in the establishment) in connection with economic issues derived from the AFiD-Panel Industrial Units. In addition analyses concerning product-related consumption of water resp. resources will be possible.

The *"AFiD-Modul Environmental Protection Expenditures"* can be combined with the AFiD-Panel Industrial Units or with the AFiD-Panel Industrial Enterprises. This module provides the option to examine the amount, stucture and developments of environmentally related investments linked to the information from the panel data sets. By merging the data on establishment- resp. enterprise level studies of the economic demand for environmental protection commodities can be conducted with regard to numerous economic items.

On the other hand, the *"AFiD-Module Environmental Protection Commodities"* enables the scientific community to take a look at the supply side of goods related to environmental protection by the industrial economy. Since the module covers details to goods, construction works and services for environment protection purposes, it can again be used with the AFiD-Panel Industrial Units to analyse a multitude of economic issues.

4. Future prospects

The boundaries for developing new paneldata or modules are not pushed yet, as an example also the development of a longitudinal data set of the corporation tax statistics is conceivable.

During the next step especially the business-related data will be combined with the panel data of the URS. Since the german business register includes only few metric variables, this will quite enlarge its information potential.

Furthermore sufficient metadata for the highly complex created data sets have to be compiled by the RDC. For adequate work with the data the scientific community requires information about the process of data-merging and the thereby conducted selections as well as information about the definitions of alike variables coming from different surveys and other methodological issues (e. g. creation of expansion factors).

Within the AFiD-Project the research potential of firm data will be considerably increased by the described data combinations: merging of unit- and enterprise data is of special interest for scientific projects, because not only more information than before is provided, but due to the surveys covering different economic activities also firms that change their line of business can be analysed. Also research on all enterprises in Germany could be conducted with the new data sets.

A following step is the combination of microdata not only from official statistics. In the pilot project "Combined Firm Data for Germany (KombiFiD)"[11] firm data of official statistics is to be merged with information from the german employment agency and the German Central Bank (Bender et al 2007) to create a complex firm data set.

[11] The acronym stands for the german *"Kombi*nierte *Fi*rmendaten für *D*eutschland".

Appendix

Table 1: **Overview of all AFiD-Products**

AFiD-Panel	German Classification of Economic Activities (WZ03)		Included surveys	Years	Characteristics	Enhancement by AFiD-Module ...
	Section	Description				
Agriculture	A	Agriculture, hunting and forestry	– Census of agriculture – Census of agricultural structure	1999, 2003, 2007	Structure and capacities of production, land use, employees	
Industrial Units	C D	Mining and quarrying Manufacturing	– Monthly report – Investment survey – Quarterly production survey – Survey of small units	1995 – 2006	Employees, turnover (domestic and non-domestic), investments, wages and salaries, sale production value	– Use of energy – Water supply and waste water disposal – Environmental protection expenditure – Environmental protection commodities – Earnings
Industrial Enterprises			– Annual report – Investment survey – Cost structure survey	2001 – 2006	Employees, turnover, investments, wages and salaries, material and commodities received, costs by type of costs, gross value added	– Environmental protection expenditure – Earnings (with limitations)
Energy Units	E	Electricity, gas and water supply	– Monthly report – Investment survey	2003 – 2006	Employees, wages and salaries, investments	
Energy Enterprises			– Cost structure survey – Investment survey	2003 – 2006	Employees, turnover, material and commodities received, costs by type of costs, taxes / subsidies	
Trade	WZ 50 WZ 51 WZ 52	Wholesale and retail trade, repair of motor vehicles, motorcycles and personal and household goods	– Annual survey in trade	1999 – 2006	Employees, turnover, wages and salaries, stock / commodities received	
Hotel and Restaurant Industry	H	Hotels and Restaurants	– Annual survey in hotels and restaurants	1999 – 2006	Employees, turnover, wages and salaries	
Services	I K	Transport, storage and communication Real estate, renting and business activities	– Structural survey in the services sector	2003 – 2006	Employees, wages and salaries, turnover, taxes / subsidies	
Business Register	D – K M – O	All units and enterprises contributing to the gross domestic product	– German Business Register	2004 – 2007	Economic activity, employees subject to social insurance contributions, turnover subject to taxation, corporate affiliation	

Literature

Bender, S. / Wagner, J. / *Zwick,* M. (2007): KombiFiD – Kombinierte Firmendaten für Deutschland, Working Paper No. 60, Leuphana Universität Lüneburg, Lüneburg.

Brandt, M. / *Oberschachtsiek,* D. / *Pohl,* R. (2008): Neue Datenangebote in den Forschungsdatenzentren. Betriebs- und Unternehmensdaten im Längsschnitt, AStA Wirtschafts- und Sozialstatistisches Archiv 3, 193 – 207.

Fritsch, M. / *Görzig,* B. / *Hennchen,* O. / *Stephan,* A. (2004): Cost Structure Surveys in Germany, Schmollers Jahrbuch / Journal of Applied Social Science Studies 124, 4, 557 – 566.

Görzig, B. / *Bömermann,* H. / *Pohl,* R. (2005): Produktdiversifizierung und Unternehmenserfolg: Nutzung der Forschungsdatenzentren der Statistischen Ämter, Allgemeines statistisches Archiv 89 (3), 337 – 352.

Görzig, B. / *Gornig,* M. (2005): Neue Produkte – Markterfolg – Entlohnung, Schmollers Jahrbuch – Journal of Applied Social Science Studies 125 (4), 509 – 524.

Görzig, B. / *Gornig,* M. / *Werwatz,* A. (2007): Produktdiversifizierung: Konvergenz zwischen ost- und westdeutschen Unternehmen: Eine Dekomposition mit Mikrodaten der amtlichen Statistik, Jahrbücher für Nationalökonomie und Statistik 227, 2, 168 – 186.

Hafner, H.-P. / *Lenz,* R. (2007): Die Gehalts- und Lohnstrukturerhebung: Methodik, Datenzugang und Forschungspotential, FDZ-Arbeitspapier Nr. 18, Statistische Ämter des Bundes und der Länder, Wiesbaden.

Kaiser, U. / *Wagner,* J. (2008): Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten, Perspektiven der Wirtschaftspolitik 9 (3), 329 – 349.

Konold, M. (2007): New possibilities for economic research through integration of establishment-level panel data of German official statistics, Schmollers Jahrbuch – Journal of Applied Social Science Studies 127 (2), 321 – 334.

Malchin, A. / *Pohl,* R. (2007): Firmendaten der amtlichen Statistik – Datenzugang und neue Entwicklungen im Forschungsdatenzentrum, Vierteljahrsheft des DIW Berlin 3 / 2007, 8 – 16.

Statistisches Bundesamt (2005a): Investitionserhebung bei Unternehmen und Betrieben des Verarbeitenden Gewerbes sowie des Bergbaus und der Gewinnung von Steinen und Erden, Qualitätsbericht.

Statistisches Bundesamt (2005b). Kostenstrukturerhebung im Verarbeitenden Gewerbe, im Bergbau sowie in der Gewinnung von Steinen und Erden, Qualitätsbericht.

Statistisches Bundesamt (2005c): Monatsbericht für Betriebe des Verarbeitenden Gewerbes sowie des Bergbaus und der Gewinnung von Steinen und Erden, Qualitätsbericht.

Statistisches Bundesamt (2005d): Produktionserhebungen, Qualitätsbericht.

Statistisches Bundesamt (2006): Unternehmensregister – System 95, Qualitätsbericht.

Statistisches Bundesamt (2008): Strukturerhebung im Dienstleistungsbereich 2006, Qualitätsbericht.

Sturm, R. / *Tümmler,* T. (2006): Das statistische Unternehmensregister – Entwicklungsstand und Perspektiven, Wirtschaft und Statistik 10, 1021 – 1036.

Verordnung (EWG) Nr. 2186 / 93 des Rates vom 22. Juli 1993 über die innergemeinschaftliche Koordinierung des Aufbaus von Unternehmensregistern für statistische Verwendungszwecke.

Vogel, A. (2009): The German Business Services Statistics Panel 2003 to 2007, Schmollers Jahrbuch / Journal of Applied Social Science Studies 129, (3), 515 – 522.

Wagner, J. (2008): Die Forschungspotentiale der Betriebspaneldaten des Monatsberichtes im Verarbeitenden Gewerbe, ASta – Wirtschafts- und sozialstatistisches Archiv 2 (3), 209 – 221.

Zühlke, S. / *Christians,* H. / *Cramer,* K. (2007): Das Forschungsdatenzentrum der Statistischen Landesämter – eine Serviceeinrichtung für die Wissenschaft, AStA Wirtschafts- und Sozialstatistisches Archiv 3 – 4, 169 – 178.