

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>.

The Research Data Center Ruhr at the RWI (FDZ Ruhr im RWI)

By Thomas Bauer, Rüdiger Budde, and Sandra Schaffner

1. Introduction

Research institutes, such as the RWI in Essen, collect a multitude of data, which is often lost for the scientific community after the respective projects are finished. Following the recommendations of the “Kommission zur Verbesserung der informationellen Infrastruktur” (Commission to improve the informational infrastructure by co-operation of the scientific community and official statistics; KVI, 2001), the RWI started in 2011 to establish a research data center – the FDZ Ruhr – in order to facilitate the use of its data by external researchers. The creation of the FDZ Ruhr is partly financed by the Federal Ministry of Economics and Technology as well as the Ministry of Innovation, Science and Research of North Rhine-Westphalia.

The basic goal of the FDZ Ruhr is to provide external researchers access to data that has been collected by the RWI either within research projects financed by its basic research budget or in the course of commissioned research projects, typically from German ministries. In order to achieve this goal, the FDZ has contacted the principals of past projects to get the permission to provide access to data collected in the respective research projects. In addition to the documentation, preparation, provision and support in the proper use of these data sets, the members of the FDZ Ruhr are working on conceptual problems concerning (1) the anonymization of microdata, the internet-based distribution of data as well as methodological issues concerning (2) the use of data obtained from the internet and (3) the generation, provision and use of geo-referenced data and the link of this type of data with other individual data sets. Data access is provided only to persons and institutions with an explicit scientific or public interest in the use of the available data. Commercial inquiries are generally not assisted.

This paper provides a brief overview of the data sets that are currently available at the FDZ Ruhr, differentiating between surveys collected in the course of research projects at the RWI and regional and geo-referenced data, the latter constituting a special focus of the Research Data Center. The paper concludes with a short outline about planned developments. Further information, including the terms and conditions of use, is available on the website of the FDZ Ruhr (<http://fdz.rwi-essen.de>).

2. Available Survey Data at the FDZ Ruhr

Table 1 summarizes the data currently available at the FDZ for use by external scientists, which cover both, individual and firm data sets. Most of them are cross-sectional data with a specific research question in mind, depending on the particular project they have been collected for. However, all of these data sets provide sufficient information to answer research questions that are beyond the specific purpose they have been collected for.

The survey *Young Families* has been collected as part of the evaluation of the effects of the German *Elterngeld* (parents' money) reform, which the RWI conducted for the German government (Kluve/Tamm, 2012; RWI, 2009). The objective of the survey was to obtain a detailed description on the behavior of parents with respect to their entitlement to parental benefits, the duration and scheduling of career breaks and returning to work as well as the usage of different types of childcare. For the first cross-section, a random sample was drawn in 2008 from the population of parents whose child was born in the first quarter of 2007 or last quarter of 2006. This sampling procedure allows comparing parents with children born shortly before and after the introduction of the *Elterngeld* on January 1, 2007. The data provides information on mother's and

Table 1
Data Sets currently available at the FDZ Ruhr

Name	Description	Unit of Observation	Data Structure
Young Families	Survey of Families to evaluate effects of parents' money	Individual	Repeated Cross-Sections collected in 2008 and 2009
Energy Panel	Survey of about 6500 households concerning their use of energy, housing and living conditions.	Households	Panel data covering the period from 2003–2010
Neighborhood Impulses	Survey of firms concerning the regional economic climate and cooperation	Firms	Cross-Section collected in 2007 and 2008
Minimum Wages and Smoking Ban	Survey of Managers of 800 Firms concerning their perceptions of the effects of the introduction of a minimum wage and a smoking ban in restaurants	Managers of Firms	Cross-Section collected in 2008
Innovative behavior of small firms	Survey of 3000 handcraft and 1000 non-handcraft firms concerning their innovation strategy.	Firms	Cross-Section collected in 2003

father's age, nationality, education, labour force status, childcare and family planning. The second data set consists of a random sample of parents with children born in April 2007 collected in 2009. In order to ensure comparability, the questionnaire is very similar to the questionnaire in 2008. It, however, additionally includes information on the role of measures at the workplace that increase the compatibility of family and working life.

The goal of the *Energy Panel* is to collect information on the energy consumption of private households. On behalf of the German Ministry of Economics and Technology, RWI together with forsa conducted surveys obtaining some 7,000 observations for each of the years spanning 2003 through 2010 (RWI und forsa, 2011, 2012). The surveys collect invoice data from households on various dimensions of their energy consumption, including quantities consumed, total costs incurred, and unit prices for electricity, natural gas, district heating and heating oil purchases. Whenever possible, participating households were also asked to record meter readings over particular time intervals to allow for data validation. For fuels usually purchased at a store, the survey participants are asked to estimate their consumption. These fuels include logs, wood briquettes, wood pellets, coal briquettes, and liquefied pressurized gas. The sur-

vey also covers information on renewable energy, including photovoltaics or solar thermal. Moreover, the data provides substantial information on socio-economic characteristics, including household size, income, ownership status of the dwelling, and the location of the household. In addition, housing characteristics are available, such as the age and the type of the house, as well as the size in square meters. The papers by Grösche/Vance (2009) and Grösche/Schröder (2010) provide an impression of the research potential of the data.

At the moment, the data from the different survey waves are processed to be available in the form of a panel. The usability of the dataset will be enhanced further by integrating it in *PanelWhiz* available for the statistical software STATA, which allows the user to select individual variables of interest (Haisken-DeNew/Hahn, 2010).

The data set *Neighborhood Impulses* has been collected as part of the Experimental Housing and Urban Development (ExWoSt) research project “Quartiers-Impulse” commissioned by the Federal Office for Building and Regional Planning (BBR) (Neumann/Jakubowski, 2006). In the project, measures to foster the local business environment were evaluated over a two-year period (2006–2008) in selected German cities. The data set comprises the results from the project- and reference areas of five cities (Bayreuth, Braunschweig, Hannover, Karlsruhe, Saarbrücken) collected in 2007 and 2008. The survey topics comprised, among others, important regional factors for firms, inter-firm cooperation and networking. A more detailed description of the data set is provided by Borne et al. (2008) and BBSR im BBR (2009).

Collected by a German-wide telephone survey in spring 2008 by the RWI and forsa, the dataset *Minimum Wages and Smoking Ban* provides information facilitated by managers of eight industries in Germany, which were likely candidates for the introduction of a minimum wage at the time the survey. These industries belong to trade or services and include hardware stores (representing wholesale trade), men’s outfitters and florists (both retailing), motor mechanics (repairs), restaurants (the hotel and restaurant industry), hairdressers (personal services), security firms (business services), and plumbing (construction). The primary research question based on this dataset has been on the employment and fiscal effects of the minimum wage (Bachmann et al., 2008). A second paper that uses this survey analyses whether employers support the introduction of industry-specific minimum wages as a cost-raising strategy in order to deter market entry. Following these research questions, managers were asked about their appraisal and potential consequences of an introduction of minimum wages. The questionnaire, asked to 800 managers, further contains information about employee turnover, employment structure, wages, zip codes etc. Managers of restaurants have further been asked about their expectations on the effects of an introduction of smoking bans, which have also been an issue of public discussion at the time the survey has been collected.

The data *Innovative Behavior of Small Firms* has been collected by RWI to evaluate the determinants of structural changes in German handcraft and non-handcraft firms, with the latter acting as a reference group. The project itself was commissioned by the Federal Ministry of Economics and Technology in 2002 to analyze the impact of technological, demographic, organizational and institutional changes on firm structure and identity (Rothgang, M. und L. Trettin, 2005; RWI, 2004). The data provides a wide range of information on various aspects of the situation of the firms in the sample, putting a special emphasis on the firm's spatial location, the structure of the employees, innovative efforts in production and the spatial range of its turnover.

3. Regional Data at the FDZ Ruhr

A major research focus of the RWI has always been on regional economic issues. Due to this focus, researchers at the RWI have a special expertise in the collection and analysis of regional data. One of the main goals of the FDZ Ruhr is to provide external scientists access to its collection of regional statistics, in particular its portfolio of highly disaggregated spatial data, and to its expertise in handling this type of data.

The majority of official regional statistics is collected with the administrative borders of political agencies in mind. Hence, the regional data usually available is typically not disaggregated at levels lower than independent cities and counties. For specific variables, data is available on the level of municipalities. However, even at the level of municipalities data is often still too highly aggregated for the specific research question of interest or, due to problems associated with political and institutional boundaries, may result in a spatial bias of the empirical results. Therefore, one of the central goals of the FDZ Ruhr is to provide data at very small spatial units of observation, to support researchers in the analysis of this type of data, to test possibilities to merge this type of data with individual survey data and to develop solutions for methodological problems arising when analyzing this type of data.

Data protection legislation puts severe restrictions on the provision and use of individual geo-referenced data. In cases where these data are available but cannot be published or shared, however, an aggregation of geo-referenced individual data may be made available at a specific spatial level (grid). In cooperation with the Research Data Center of the Federal Employment Service in the Institute for Employment Research and the DIW Berlin, researchers at the FDZ Ruhr are currently working on the project "Nachbarschaftseffekte: Die Analyse individuell-rationalen Verhaltens im sozialen Kontext (Neighborhood Effects)" financed by the *Pakt für Forschung und Innovation* of the German Government. Part of this project is to geo-reference the data available at the Research Data Center of the Federal Employment Service in the Institute for Employ-

ment Research and to merge this data and the SOEP, which is already geo-referenced, with aggregated data at small spatial units of observation, i.e. at grids of the size of one squared kilometer.

To facilitate the exchange of data with other European researchers, the grid-level data in this project, as well as other small spatial unit data available at the FDZ Ruhr, follow the guidelines of the INSPIRE Directive of the European Union. (D2.8.I.2 INSPIRE Specification on *Geographical Grid Systems – Guidelines*). In its report on geo-referencing data, the RatSWD states that the above mentioned specification on geographical grids is mandatory for Germany (RatSWD, 2012, 56). Some European nations already offer their demographic data on the level of these grids via EUROSTAT. We adhere to the proposed spatial notation of the European Forum for Geostatistics.

The following sketch demonstrates some of the challenges in handling this type of regional data. As mentioned before, most regional data are related to administrative boundaries. Researchers working with this type of data have to take into account that these borders vary over time. With grid data, however, each region has got the same size (i.e. 1 sq. km) (see Figure 1).

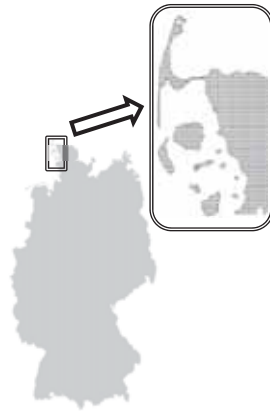
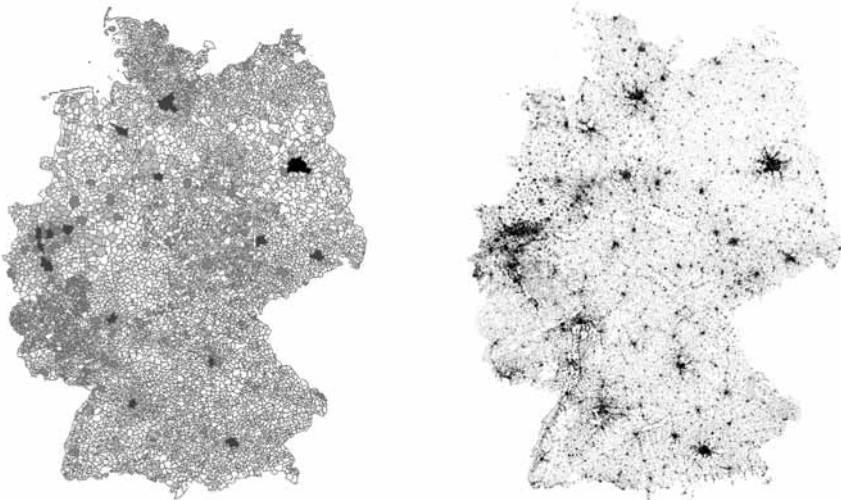


Figure 1: Schematic Approach: Reporting Grid for Germany

The position of each grid is unambiguously defined and only differs by the amount of its attributes. Figure 2 gives an example how the population is locally distributed within Germany. The left map represents the level on the German administrative level “Gemeinde” while the right map shows the same data on grid-level. The figure makes apparent, that the regional distribution of the population obtained by the official data, which is aggregated at the level of municipalities, does not provide an adequate picture of the true regional distribution.

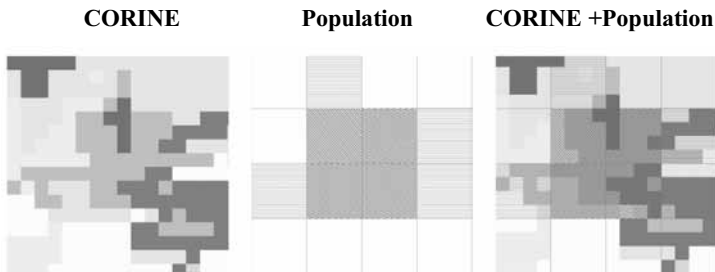
Population, "Gemeinde"-level, 2010

Population, 1 sq. km grid level, 2010



Source: GENESIS-online Regional, microm GmbH.

Figure 2: Regional distribution of the German population of different regional levels, 2010



Source: European Environment Agency, microm GmbH.

Figure 3: Regional Matching CORINE-Land Cover Data, 2006, with Population Grid-Data, 2010

As there are various ways to lay a grid over a surface, it is important that all grids are designed in the same way, i.e. that the borders of the squares are always the same. Data from different sources that use the same grid definition can be easily combined. Figure 3 shows an example where data from the "CORINE Land Cover Raster Data" of the EU has been matched with population data from the FDZ Ruhr. The left figure shows the share of sports and leisure activity areas in a region obtained from the CORINE data, which is

based on a grid of 250 meters times 250 meters. The figure in the middle shows the size of the population in the same region based on a grid of 1 kilometer times 1 kilometer, which is available at the FDZ Ruhr. Both data sets can be combined very easily to study the availability of sports and leisure activity available to the population in the data of the FDZ Ruhr, since combining both grids does not lead to any regional overlaps.

In addition to the spatial data generated in the above mentioned project on neighborhood effects, the FDZ Ruhr currently provides access to the following variables for the year 2010 at the 1 sq. km grid level, which have been collected by the FDZ Ruhr in cooperation with the microm GmbH: buildings by type of use (commercial/residential), residential buildings by type and number of households, households by different categories (including ethnicity), total population by age group and sex, indicators for cars with respect to brand density and segments, purchasing power, unemployment, and household credit ratings. Interested researchers are able to obtain individualized maps, raster data in the form of a Scientific Use File (minimum of 20 grids) or may use the data at the FDZ Ruhr in Essen or (in the future) via remote access. Merging and matching these datasets with other user-specific datasets is in principle allowed.

Moreover, the FDZ Ruhr is geo-referencing available data and matches them to the aforementioned raster projections. Such a procedure makes it possible, for example, to match information from ImmobilienScout24 on real estate prices, rents and housing characteristics (Bauer et al., 2013) with individual datasets. Such combined data sets can be used to analyze a vast number of new research questions. Bauer, Fertig and Vorell (2011), for example, use this type of merged data to identify neighborhood effects in the occurrence of unemployment. The management of privacy concerns that arise when one aims to merge different data sets are facilitated by a standardized procedure.

4. Planned Activities in the Coming Years

In the coming years, it is intended to enlarge the available micro-data at the FDZ Ruhr. Currently the team is preparing a panel data set of important German stock corporations, which covers important information of these firms published in their business reports, and a cross-sectional data set of slightly more than 4000 individuals, which has been collected in order to analyze the regional identity of the population in the Ruhr area. It is also working on the preparation and provision of different internet-based data sets, such as the data from ImmobilienScout24, for external researchers. In addition, the FDZ Ruhr is digitalizing the German theatre statistic, which is yet mostly only available on paper. The latter must be seen in the attempts of the FDZ Ruhr, to continuously expand its' portfolio of regional economic data, in particular grid-level data.

As the FDZ Ruhr is still in its start-up phase, it is necessary to adjust its infrastructure and its regards to RWI contents and public demands. In the coming years, the FDZ Ruhr aims to enlarge the available office space and technical infrastructure for external researchers, which includes an expansion of the data security zone in order to enable more external researchers to work on confidential data with established statistical software packages as well as with geographical information systems.

References

- Bachmann, R./Bauer, T. K./Kluve, J./Schaffner, S./Schmidt, C. M.* (2008): Mindestlöhne in Deutschland – Beschäftigungswirkungen und fiskalische Effekte, RWI Materialien 43, Essen.
- Bachmann, R./Bauer, T. K./Kröger, H.* (2012): Minimum Wages as a Barrier to Entry – Evidence from Germany, Ruhr Economic Papers No. 329, RWI, RUB.
- Bauer, T. K./Fertig, M./Vorell, M.* (2011): Neighborhood Effects and Individual Unemployment. Ruhr Economic Papers No. 285, RWI, RUB.
- Bauer, T. K./Feuerschütte, S./Kiefer, M./an de Meulen, P./Micheli, M./Schmidt, T./Wilke, L.* (2013): Ein hedonischer Immobilienpreisindex auf Basis von Internetdaten: 2007–2011, AStA – Wirtschafts- und Sozialstatistisches Archiv 7, 5–30. DOI: 10.1007/s 11943-012-0125-7.
- BBSR im BBR* (2009): Quartiers-Impulse – Neue Wege zur Stärkung der lokalen Wirtschaft, Bonn.
- Borne, M./Geißler, H./Gutzmer, M./Jakubowski, P./Neumann, U./Scheuer, M.* (2008): Quartiers-Impulse. Neue Wege zur Stärkung der lokalen Wirtschaft. Ein ExWoSt-Forschungsfeld, ExWoSt-Informationen 31 (4).
- Commission for the Improvement of the Informational Infrastructure between Research and Statistics* (2001): Towards a better informational infrastructure, Baden-Baden.
- Grösche, P./Schröder, C.* (2011): Eliciting Public Support for Greening the Electricity Mix Using Random Parameter Techniques, Energy Economics 33 (2), 363–370.
- Grösche, P./Vance, C.* (2009): Willingness-to-Pay for Energy Conservation and Free-Ridership on Subsidization – Evidence from Germany, Energy Journal 30 (2), 135–154.
- Haiskan-DeNew, J./Hahn, M. H.* (2010): PanelWhiz: Efficient Data Extraction of Complex Panel Data Sets – An Example Using the German SOEP, Schmollers Jahrbuch 130 (2010), 643–654.
- Kluve, J./Tamm, M.* (2012): Parental Leave Regulations, Mothers’ Labor Force Attachment and Fathers’ Childcare Involvement: Evidence from a Natural Experiment, Journal of Population Economics 26 (3), 983–1005. DOI: 10.1007/s 00148-012-0404-1.
- Neumann, U./Jakubowski, P.* (2006): Quartiers-Impulse: Neue Wege zur Stärkung der lokalen Wirtschaft. Ein neues Forschungsfeld im Experimentellen Wohnungs- und Städtebau (ExWoSt) des Bundes, Werkstatt: Praxis (45), 61–71.

Rat für Sozial- und Wirtschaftsdaten (2012): Georeferenzierung von Daten. Situation und Zukunft der Geodatenlandschaft in Deutschland, Berlin.

Rothgang, M./Trettin, L. (2005): Patterns and Determinants of Innovation in the German Craft Sector. In: Würth, R./W. Gaul/V. Jung (eds.), *The Entrepreneurship – Innovation – Marketing Interface*, Künzelsau, 365–386.

RWI (2004): Determinanten des Strukturwandels im deutschen Handwerk – Endbericht, RWI Projektberichte, Essen.

RWI (2009): Evaluation des Gesetzes zum Elterngeld und zur Elternzeit – Studie zu den Auswirkungen des BEEG auf den Arbeitsmarkt und die Vereinbarkeitsplanung – Endbericht, RWI Projektberichte, Essen.

RWI and forsa (2011): Erhebung des Energieverbrauchs der privaten Haushalte für die Jahre 2006–2008, RWI Projektberichte, Essen.

RWI and forsa (2012): Erhebung des Energieverbrauchs der privaten Haushalte für die Jahre 2009–2010.