

Improvement of Access to Data from Official Statistics – the case of Germany

**Bender Stefan¹, Himmelreicher Ralf², Zühlke Sylvia³,
Zwick Markus⁴**

¹Institute for Employment Research, e-mail: stefan.bender@iab.de

²German Federal Pension Insurance, e-mail: ralf.himmelreicher@drv-bund.de

³ Statistical Office of Northrhine-Westphalia, e-mail:
sylvia.zuehlke@it.rw.de

⁴Federal Statistical Office, e-mail: markus.zwick@destatis.de

Abstract

The four publicly funded research data centres in Germany – the Research Data Centre (FDZ) of the Federal Employment Agency of Institute for Employment Research, the FDZ of the German Pension Insurance, the FDZ of the Statistical Offices of the Länder and the FDZ of the German Federal Statistical Office – have made a significant improvement to the data and services available to researchers over the past few years. Their services are widely used, empirical findings lead to refereed publications and the state of research in rendering microdata anonymous has made great leaps. Many policy decisions are now planned and evaluated on the basis of data originating from the FDZ. Germany has gone from the bottom of Europe's league with regard to the use of individual data to an innovative provider of new ideas, such as on access to individual data for teaching purposes and linked employer-employee datasets. In 2007, the FDZ developed criteria for their specific design in conjunction with the German Council for Social and Economic Data. The topics of the paper are to describe the key criteria representing a common working basis for the FDZ, detailed descriptions of the four FDZ and an outlook over future German developments.

Keywords: Research Data Centre, Data Access, Data Protection

1. Introduction

The four publicly funded research data centres (FDZ) in Germany – the Research Data Centre of the Federal Employment Agency at the Institute for Employment Research (FDZ-BA), the Research Data Centre of the German Pension Insurance (FDZ-RV), the Research Data Centre of the Statistical Offices of the Länder (FDZ-SOL) and the Research Data Centre of the German Federal Statistical Office (FDZ-FSO) – have made a significant improvement to the data and services available to researchers over the past few years.¹ Founded on a recommendation by the German Commission on Improving the Informational Infrastructure between Science and Statistics (KVI 2001) and funded in the project phases by the German Federal Ministry of Education and Research (BMBF), the centres have developed in a form that was not initially anticipated. Their services are widely used, empirical findings lead to refereed publications and the state of research in rendering microdata anonymous has made great leaps. Germany has gone from the bottom of Europe's league with regard to the use of individual data produced by empirical research with public funding, to an innovative provider of new ideas, such as on access to individual data for teaching purposes and linked employer-employee datasets.

2. The German Council for Social and Economic Data - Criteria for Research Data Centres

Research data centres are institutions with the main purpose of providing simple, transparent and high-quality access to microdata suitable for statistical analysis, while maintaining data protection and data security. Moreover, the research data centres are intended to

¹ Two data service centres – the German Microdata Lab (GML) at ZUMA and the International Data Service Center at IZA (IDSC) – were also set up as part of this development, and have also worked very successfully (Schneider/Wolf 2008).

contribute to improving cooperation between the data users from the scientific community and the respective data producers. The research data centres are thus an interface between the data producers' supply of data and the demand for these data from the science side. Strictly observing data protection regulations, they enable the following individual data access: anonymous microdata files, controlled remote data access and workplaces for guest researchers in the research data centres.

In 2007, the research data centres developed criteria for the specific design of research data centres in Germany in conjunction with the German Council for Social and Economic Data (RatSWD):²

1. The data made available to the scientific community arise for statistical purposes as part of public administration processes, research and evaluation, and are produced using public funding.
2. Access to the data is subject to the legal provisions of data protection and data security in the specific area. The task of the research data centres is to provide easier access through specific regulations.
3. Access to microdata is subject to legal regulations guaranteeing equal treatment of data users. Correspondingly, the research data centres ensure transparent and standardised access regulations and do not undertake any evaluation of the content of the research projects applying for access, merely checking for data protection or contractual permissibility.
4. As well as providing access possibilities, the research data centres also produce data products for easy analysis and comprehensive data documentations. The research data centres' tasks also include organising and participating in academic events in order to present the available data material, and presenting the available data and access to potentially interested parties.
5. The work within the research data centres must not be restricted to service activities, as this would ultimately be equivalent to an exit from the scientific stage. Scientific research within the research data centres enables access to further skills and qualifications and

² Long version available under http://www.ratswd.de/download/publikationen_rat/RatSWD_FDZKriterien.PDF.

participation in scientific events, and also the publication of own work in the relevant journals.

6. Research within the research data centres is not coupled with exclusive access for the data producers. The anonymous microdata is made available simultaneously to all researchers, at least via controlled remote data access or at workplaces for guest researchers.

3. The Four Public Funded Research Data Centres

There are currently four research data centres in Germany recommended by the RatSWD, which are described below.

3.1 Research Data Centre of the German Federal Statistical Office (FDZ-FSO)

The FDZ-FSO was the first research data centre to be set up on the recommendation of the KVI in 2001 (see Zühlke et al 2004). The core activity of the FDZ-FSO, processing user requests, is now funded completely from original in-house sources. The FDZ-FSO also receives funding from the BMBF within scientific projects.

The most important official statistics are now available in the FDZ-FSO and the Statistical Offices of the Länder, as a joint service. Information on data and forms of access is available at www.forschungsdatenzentrum.de.

The most intensively used form of data is the SUFs. Approximately 888 standardised SUFs have been provided for 387 different projects from mid 2004 to 2008, when the FDZ-FSO was first evaluated. Overall, however, demand for standardised SUFs is stagnating, whereas demand for individual datasets at the workplaces for guest researchers and via controlled remote data access is increasing. 41 researchers have visited the research data centre since 2004, with controlled remote data access used in 82 projects. 24 further projects are currently taking place using the two forms of access.

The *RDC Working Papers* series was initiated to present the wide use of the official microdata. To date, nearly thirty working papers

have been published in this series, available via the homepage. The Federal Statistical Office's book series *Statistik und Wissenschaft* also includes various volumes of articles reflecting the dialogue between the research data centre and the scientific community.

The FDZ-FSO developed the series *CAMPUS-Files* especially for teaching at university level. These files consist of absolutely anonymous microdata, allowing students to learn methodological skills and analyse sociological and economic issues. These data are available free of charge via the website (see Zwick 2007).

The FDZ-FSO's work focuses on further development of the access routes, anonymity methodology and conceptional development. In order to strengthen its anchoring in the scientific community, the research data centre is strongly present at relevant German conferences. It also offers its staff the possibility to gain PhDs using the research data centre's resources, via two-thirds contracts.

3.2 Research Data Centre of the Statistical Offices of the Länder (FDZ-SOL)

The core task of the FDZ-SOL is to provide easier access to the individual data of the statistical offices of the Länder for scientific research (Zühlke et. al.). In order to realise this task, a regional infrastructure was set up, enabling nationwide access to official microdata for the scientific community in 16 regional locations. Moreover, a centralised data administration was established, which makes inter-regional use of the microdata of the Statistical Offices of the Länder simpler.

The activities of the FDZ-SOL have led to a broad range of microdata on various subjects for the scientific community. A total of over 60 statistics are currently available for use in academic research projects from the fields of social issues, the economy, agriculture, the environment, justice administration and taxation (see www.forschungsdatenzentrum.de.)

Use of the FDZ-SOL has increased steadily since 2004, with the number of applications for use multiplying fivefold by 2008 from 31 to 157. On average, each application requires access to six different datasets, so that more than 3,600 datasets have been provided for

research purposes to date. Due to the decentralised infrastructure, most data use takes place at the workplaces for guest researchers or via controlled remote data access, and is thus particularly labour-intensive for the research data centre.

To guarantee a further user-oriented development of the range of services the FDZ-SOL has regularly carried out evaluations since mid-2005. On the whole, the results show a high users' satisfaction with the range of services.

The FDZ-SOL has been funded by the BMBF since the beginning of 2004. The evaluation of the FDZ-SOL in October 2006 was positive and recommended establishing the FDZ-SOL on a permanent basis.

3.3 The Research Data Centre of the Federal Employment Agency at the Institute for Employment Research (FDZ-BA)

The FDZ-BA (Bender et al. 2008, Kohlmann 2005) was founded in December 2003 and in 2006 it was permanently established as an independent research data centre.

The FDZ-BA prepares individual datasets developed in the sphere of social security and employment research. On its website (<http://fdz.iab.de>) data documentation and working tools such as the FDZ Datenreport and the FDZ Methodenreport are available. By means of website, workshops and users' conferences the FDZ-BA makes it easier for external researchers to work with its datasets.

The FDZ-BA serves not only the national but also the international market. One important step towards internationalisation in 2007 was releasing web pages in English and having nearly every data documentation translated. Since the cost of a stay in Nuremberg for visitors from abroad is higher than it is for locals, the FDZ-BA set up a grant in 2007.

The number of approvals for data access has increased continuously from 46 projects (in 2004) to 108 (in 2008). The number of researchers using data of the FDZ-BA raised from 48 (in 2004) to 375 (in 2008); 69 of them are working outside Germany. Up to 2008, researchers had published 347 articles, 69 of them in refereed journals.

The FDZ-BA's own research activities are well documented through its publication record. In addition, the FDZ-BA participates in a number of externally funded projects, co-financed by the German Research Foundation (DFG), the Federal Ministry of Education and Research (BMBF) or the Leibniz Association and carried out in cooperation with universities, research institutes and, of course, with the other research data centres.

3.4 Research Data Centre of the German Federal Pension Insurance (FDZ-RV)

The FDZ-RV was founded in January 2004 and is now based in Berlin and Würzburg. During its initial setup phase, the FDZ-RV was funded by the BMBF from 2004 to 2008.

The core task of the FDZ-RV is to recover the data treasures of the German federal pension insurance (Rehfeld / Mika 2006). Alongside the microdata itself, the research data centre provides methodological information and commentaries intended to help simplify analyses using data from the German federal pension insurance. Current information is available at www.fdz-rv.de.

The FDZ-RV has realised the projects it agreed to fulfil for the funding institution (BMBF). Firstly, it has established an infrastructure within the research data centre, and secondly, it has taken the research data centre from a pilot project to a permanent institution. Thirdly, the range of data and the use possibilities have been extended considerably. In both Berlin and Würzburg, the micro datasets of the federal pension insurance are processed in cooperation with the respective departments and the data users, to make them available to researchers particularly in the form of user-friendly scientific use files.

The statistics of the German Federal Pension Insurance can be divided into datasets that focus on biographical information in combination with retirement and insurance and in a special dataset for rehabilitation (Himmelreicher / Stegmann 2008). The extended data range provided free of charge by the FDZ-RV, which now also includes longitudinal data, represents a considerable improvement in usage possibilities for research. As the FDZ-RV data have now been

used in numerous scientific disciplines by more than three hundred researchers and an increasing number of presentations and publications are being written on the basis of the data, the research data centre is becoming increasingly well known in the scientific community. However, once the BMBF funding expires at the end of 2008, it will become difficult to provide the familiar standard of services to an increasing number of researchers with fewer staff.

3.5 Sustainability of the German RDC network

At the end of the phases funded by the BMBF, the research data centres are facing new challenges. The FDZ-BA has meanwhile been integrated into the Federal Employment Agency with all capacities from the funding phase, and now carries out its work as an organisational unit in their institution. The FDZ-RV has been established as a permanent institution, equipped with basic funding to meet the key infrastructural needs. Additional third-party funding has to be obtained for research projects. For the FDZ-FSO and FDZ-SOL, possibilities for establishment on a permanent basis are still under discussion.

4. The German Developments in an International Perspective

The developments in Germany are rather new and fast. For example, “Managing Statistical Confidentiality & Microdata Access” (United Nations 2007) does not contain a single word about the German situation³.

The decentralized RDC solution is a direct reaction to the decentralized data production in Germany. But unlike in the USA, the RDCs in Germany have developed criteria for the specific design of RDCs. Thereby it is guaranteed that every RDC has the same standards for data access. For researchers an easy access to all data

³ To be precise, Germany is mentioned in Annex 1.23 for joining IPUMS in the near future.

sets of the above named institutions is possible, because the RDCs ensure transparent and standardised access regulations. The decision if a request on data access is approved or not is based solely on data protection regulations. The access to the data is nearly free, because the data in Germany are treated as a “public good”. Special rates and easy access to campus files for younger researchers stimulate the scientific offspring.

The access to German data sets is not exclusive for German researchers. After the start up phase some of the RDCs/DSCs – like the FDZ-BA or the IdZA – tried to attract Non-German researchers to German data. The FDZ-BA, for example, provides grants for guest stays in Nuremberg.

The German data are opened up for researchers systematically by the RDCs. The research data centres coordinate basic issues of data access for research purposes among each other and work in close conjunction on various projects to extend the range of data available and the access routes. Further development of the data range will focus on integrating statistics in the near future. The projects “Amtliche Firmendaten für Deutschland” (official firm data for Germany – AFiD) and “Kombinierte Firmendaten für Deutschland” (combined firm data for Germany – KombiFiD) will extend the range of data in two directions: AFiD will bring together economic and environmental data from the Statistical Offices by means of the German Company Register (URS) on the microdata level. KombiFiD goes one step further, uniting company data across the boundaries of the individual data producers as part of a feasibility study on a joint dataset (Hethey/Spengler 2008). In addition, processes for statistical matching of survey and process-produced data are being tested, e.g. between longitudinal data of the Research Data Centre of the German Federal Pension Insurance and the DIW’s SOEP (Rasner et al. 2007) or the IAB’s longitudinal data or in several of the IAB’s own projects (Bender et al. 2008b, to quote one example).

Although data access in Germany is well developed, the RDCs will move to an easier access in the near future. Comparable to the Canadian situation the FDZ-SOL has more than 16 local units all over Germany where access to the data is available. The German RDCs are working on an RDC in RDC solution, which would give researchers the opportunity to have access to all available German data sets of

different producers in every RDC.

Remote data access has already been put to successful use in other European countries. The researchers are provided with direct access to the microdata at a specially set-up workplace in their own institutions via a secure internet connection. The research data centres are currently addressing the issue of requirements for introducing this access route in Germany, but have to find technical solutions in the German legal requirements.

Regarding the provision and usage of micro data produced with public funding, Germany has moved from the bottom of Europe's league to an innovative provider of new ideas, such as providing access to individual data for teaching purposes and to linked employer-employee datasets even for Non-Germans.

References

- Bender S. et al. (2008a): FDZ Annual Report 2007, *FDZ Methodenreport*, 02/2008, Nuremberg, 30 p
- Bender S. et al. (2008b): WeLL - Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens * Projektbericht, *FDZ Methodenreport*, 05/2008 (de), Nuremberg, 37 p
- Hethey, T., Spengler, A. (2008): *Matching Process generated Business Data and Survey Data - The Case of the KombiFiD in Germany*, mimeo, Nuremberg
- Himmelreicher, R., Stegmann, M. (2008): New possibilities for socio-economic research through longitudinal data from the Research Data Centre of the German Federal Pension Insurance (FDZ-RV), *Schmollers Jahrbuch*, 128/4, (forthcoming)
- Kohlmann, A. (2005): The Research Data Centre of the Federal Employment Service in the Institute for Employment Research, *Schmollers Jahrbuch*, 125, 437-447
- KVI – Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (2001): *Wege zu einer besseren informationellen Infrastruktur*, Wiesbaden
- Rasner A. et al. (2007): Best of Both Worlds: Preparatory Steps in Matching Survey Data with Administrative Pension Records - The

- Case of the German Socio-Economic Panel and the Scientific Use File Completed Insurance Biographies 2004. *SOEP Data Documentation*. Deutsches Institut für Wirtschaftsforschung, Berlin
- Rehfeld, U., Mika, T. (2006): The Research Data Centre of the German Statutory Pension Insurance (FDZ-RV), *Schmollers Jahrbuch*, 126/1, 122-127
- Schneider, H., Wolf, C. (2008): Die Datenservicezentren als Teil der informationellen Infrastruktur, Rolf, G. / Zwick, M. / Wagner, G. (Ed.): *Fortschritte der informationellen Infrastruktur in Deutschland*, Baden-Baden, 236-249
- United Nations (2007): Managing Statistical Confidentiality & Microdata Access – Principles and Guidelines of Good Practice, New York and Geneva
- Zühlke, S., Zwick, M., Scharnhorst, S., Wende, T., (2004): The research data centres of the Federal Statistical Office and the statistical offices of the Länder, *Schmollers Jahrbuch*, 4, 567-578
- Zwick, M. (2007): CAMPUS-Files, Free Public Use Files for Teaching Purposes, *Journal of Applied Social Science Studies*, 127, 655-668