

European Data Watch

This section will offer descriptions as well as discussions of data sources that may be 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 will start with data from German speaking countries that allow international comparative research. While most of the data will be at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) will be included, too. Suggestions for data sources to be described in future columns (or comments on past columns) should be send to: Joachim Wagner, University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to <wagner@uni-lueneburg.de>.

The Research Data Centre of the German Statutory Pension Insurance (FDZ-RV)

By Uwe G. Rehfeld and Tatjana Mika

1. Introduction: Administrative process data from public pension institutes as a basis for the Statistical Information System

Participation in the statutory pension insurance scheme is mandatory in Germany for all persons employed in the private or public sector. Additionally, contributions are paid out of unemployment insurance in the case of the unemployed, out of health insurance in the event of long-term illness, and from the state for people in military or civilian national service. The majority of the population thus comes into contact with the pensions insurance system at some point or another in life, and the pension insurance system has data on about 90 % of the entire population. The statutory old age and disability pension – due to its income replacement tasks and broad social basis – provides

the main income source after retirement. Over and above this function, the pension insurance scheme provides for rehabilitation in the event of inability to work. It therefore also includes data on health status with regard to working ability of its constituents. Since the 1960s, all data have been collected in individual administrative accounts. These are the original sources for present and retrospective statistics.

2. Centralized statistical reporting

Based on data provided by all its 26 member institutes, the Federation of German Pension Insurance Institutes compiles data and publishes several statistical reports annually (see Figure 1). These reports and some additional services are the backbone of social security planning in the fields of old-age pensions, disability pensions, and rehabilitation.

The statutory pension insurance scheme has a decade-long history of in-house statistical research on the number and structure of compulsorily insured persons, pension entitlements and rehabilitative interventions. Old age social security and savings became the focus of research and science in the 1960s. Despite the fact that these data are secured to ensure the protection of individuals' privacy, there are technical and organizational ways to make research on these data possible. In the future, the database will also be available to scientists from other non-profit institutions, provided that the research qualifies as social security research¹.

To this end, the federation began work in 2004 establishing a special service facility. The aim of the Research Data Centre of the German Statutory Pension Insurance (FDZ-RV) is to provide data on the German pension funds for scientific research. Both the research community and the Ministry for Education and Research requested that access be granted, and research allowed, on data held by government agencies, including data held by the social security administration. To fulfill this demand, the Research Data Centre for pension insurance was established and placed under the control of the Federation.

The German Statutory Pension Scheme is itself currently undergoing reform. It used to be organized along the lines of professions (blue- or white-collar employees, sailors, miners and railroad workers) and, in the case of blue-collar workers was further divided according to regions (*Landesversicherungsanstalten*). This resulted in a multitude of pension insurers, each of which collected data for their own constituency.

¹ Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (2001), *Wege zu einer besseren informationellen Infrastruktur*.

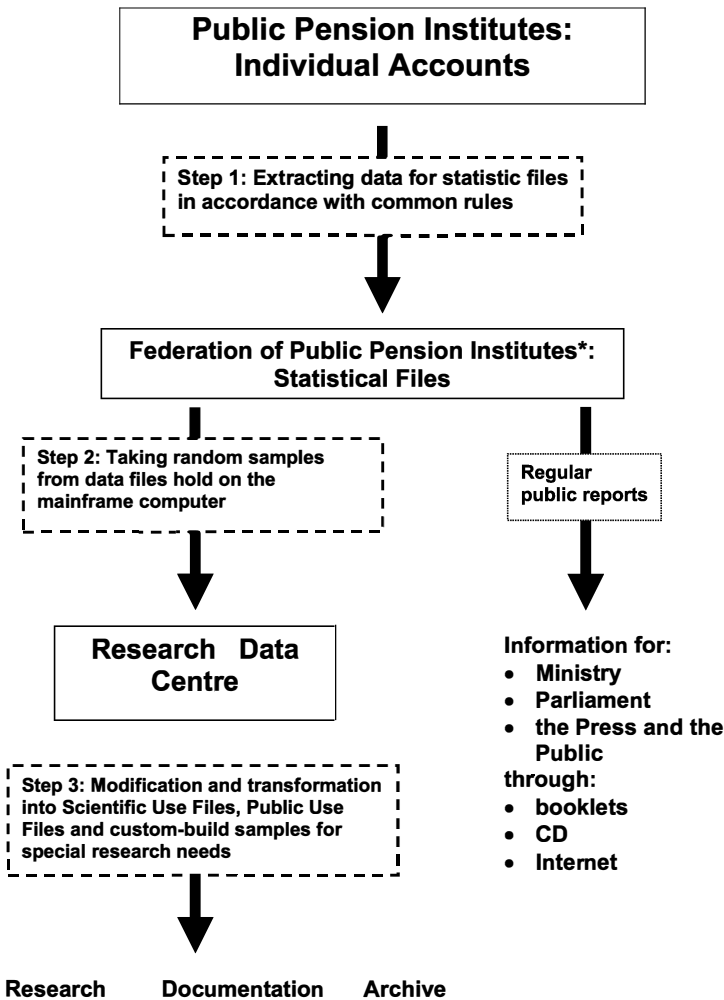


Figure 1: Data Sources and their transformation for Scientific Use

The structural reform of 2005 will change this framework by limiting the number of insurance institutes, and merging the Federation of German Pension Insurance Institutes (VDR) with the white-collar pension insurance institute (BfA) to form a new body (*Deutsche Rentenversicherung Bund*). The statistical report system² remains unchanged. The data centre of the public pension institutes in Wuerzburg will continue to archive all statistical data. The

² Rehfeld, U. G. (2001), Die Statistiken der gesetzlichen Rentenversicherung, in: DRV 1 – 2, 129.

system of statistical reporting on German statutory pensions will continue to rely on this central point of data collection. It will also be the basis for the Research Data Centre, which started with two labs in Berlin and Wuerzburg.

3. Micro Data files for the science community

Data from the statutory pension insurance can be useful in many research areas, especially economics, sociology, demographic research and other social sciences. Different data files give opportunities for research on the fields of insurance, pension payment and health as Figure 2 shows in an overview. The high quality of most variables and their comprehensive long-term nature make these data very useful. However, the data must be transformed into micro data files, which can be analyzed with statistical tools on standard PCs using programs such as SPSS, Stata or SAS. This process of transformation leads to three types of data files offered via the Research Data Centre of the Statutory Pension Fund. *Scientific Use Files* are offered as a general bases for research questions that mainly concern contributions paid and benefits drawn from the pension funds. *Public Use Files* are a tool for teaching university courses and are limited in their content, especially with regard to the number of cases, and

Insurance Statistics	Pensions statistic	Rehabilitation statistics
<u>Biographic data:</u> random sample of individual pension records (Versichertenkontenstichprobe)	<u>Biographic data:</u> Random sample of completed insurance biographies (Stichprobe Vollendete Versichertenleben)	<u>Biographic data:</u> 10 year-Retrospective social-medical observatory on medical treatment and occupational rehabilitation - occupational status - retirement information (Reha-Renten-Verlaufsstatistik)
<u>Cross sectional data:</u> Insured persons (Aktiv Versicherte)	<u>Cross sectional data:</u> Pensions in payment (Rentenbestand) Newly awarded pensions (Rentenzugang) Discontinued pensions (mostly by death) (Rentenwegfall)	<u>Cross sectional data:</u> medical treatment and occupational rehabilitation (Rehabilitationsleistungen)

Figure 2: Public Pension Micro Data Files

are therefore easy to handle. Finally, the Research Data Centre (FDZ-RV) offers data sets for special research interests (*Themendatensätze*) such as demographic data files, which have a more limited scope, which means that the files include less variables but provide more detailed information.

The first data files were handed over to researchers in February 2005. The scientific use file included information on newly granted pensions in 2003 (*Versichertenrentenzugang 2003*). The second offer is a bundle of data files for demographic research (*Themendatensatz Demografie*), which contains data from 1993, 1995, 1997, 1999, 2001 and 2003 and includes information on pensions paid in each respective year (*Rentenbestand*) as well as on discontinuation of pensions (*Rentenwegfall*).

4. Research Data Centre (FDZ-RV) under construction

The Research Data Centre was started in 2004 in order to develop micro data sets, customized for diverse research purposes. A workshop in June 2004 established a closer relationship to scientists who are doing research in this field and who have presented their data requests³ in the context of their research projects. The conference transcript is now available as *DRV Schriften* 55. It can be ordered via the Web: www.deutsche-rentenversicherung.de. The next workshop is to be held in June 2006 and the conference papers will be published soon thereafter.

From 2004 to 2006 the Research Data Centre will be in its first stage of development. During this time, the FDZ-RV is supported both on the government side by the Federal Ministry of Education and Research (BMBF) and on the Deutsche Rentenversicherung Bund side by the Research Network on Pensions (FNA). After this initial period of development, several data sets for a variety of research interests will be offered to the scientific community.

5. Tasks and challenges of the Data Centre

Data held in the German social insurance institutions are legally protected by the Social Security Law, which differentiates between data that deserve normal protection and data that deserve special protection. The data under special protection must not be given to anybody except legal enforcement institutions (e.g. the state attorney). Data with a normal protection level can be handed over for research purposes if the persons whose data are concerned and the persons responsible for data security in the pension funds institutions

³ Himmelreicher, R. (2004), Daten der Rentenversicherung und Bedarf der Wissenschaft. Befunde einer Befragung unter potentiellen Nutzern des Forschungsdatenzentrums der gesetzlichen Rentenversicherung (FDZ-RV), DRV 8.

agree to the transfer⁴. This procedure is not only costly, but also takes a long time to pass all responsible boards. As a new and quicker procedure, the Research Data Centre is now releasing anonymized Scientific Use Files.

The main task of the Research Data Centre is the conciliation of research needs and data protection. The Research Data Centre aims at providing the data much more quickly by tailoring the data files in such a way that they are successfully anonymized while offering enough information for the research question at hand⁵. The Research Data Centre is, in close cooperation with the data protection officer, developing a process of anonymization which includes the taking of random samples, the reduction of the data set regarding the number of variables, and the reduction of information on very small groups in the population, such as, for example, people holding foreign citizenship. Samples are also limited to areas with populations of more than 500,000. This concept has been designed in accordance with the anonymization procedures developed for the German Micro Census.⁶ However, the Research Data Centre also offers special data files that combine in-depth information on some characteristics, while leaving out variables included in other data files. Eventually, the research data centre will offer a broad variety of ready-made data files, which can be ordered for research projects on short notice. The main challenge for the Data Centre will be to adapt changing and developing data files of the institution and additional surveys, and to make them available for research purposes. The result is a welcome improvement of the old system.

6. International Data Access

So far, access to social security data from countries outside the European Union is forbidden by German law. Therefore, the Research Data Centre of the German Statutory Pension Insurance (FDZ-RV) is not allowed to send data abroad. Nevertheless, all persons working in Germany and doing research in the field of social security are permitted access to these data. Data are made available to accredited research institutions situated in Germany upon placing a research request, or via safe scientific work stations in the Research Data Centre in Berlin or in Wuerzburg.

⁴ Heese, C. (2004), Aspekte des Datenschutzes im Forschungsdatenzentrum des Rentenversicherung, in: Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im Aufbau, DRV Schriften.

⁵ Stegmann/Luckert/Mika (2005), Die Bereitstellung prozessproduzierter Daten der GRV im Forschungsdatenzentrum der Rentenversicherung (FDZ-RV): Grundsätze zur faktischen Anonymisierung von Mikrodaten und zu Gastwissenschaftler-Arbeitsplätzen.

⁶ Müller/Blien/Knoche/Wirth (1991), Die faktische Anonymität von Mikrodaten, Statistisches Bundesamt (Hrsg.).

References

- Heese, C.* (2004): Aspekte des Datenschutzes im Forschungsdatenzentrum des Rentenversicherung, in: Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im Aufbau, DRV Schriften, 41.
- Himmelreicher, R.* (2004): Daten der Rentenversicherung und Bedarf der Wissenschaft. Befunde einer Befragung unter potentiellen Nutzern des Forschungsdatenzentrums der gesetzlichen Rentenversicherung (FDZ-RV), DRV 8, 472
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik* (Hrsg.) (2001): Wege zu einer besseren informationellen Infrastruktur, Baden-Baden.
- Luckert, H.* (2004): Statistikdaten der gesetzlichen Rentenversicherung – ein grober Überblick, in: Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im Aufbau, DRV Schriften, 24.
- Müller, W./Blien, U./Knoche, P./Wirth, H.* (1991): Die faktische Anonymität von Mikrodaten, Band 19 der Schriftenreihe Forum der Bundesstatistik, Statistisches Bundesamt (Hrsg.), Wiesbaden.
- Rehfeld, U. G.* (2001): Die Statistiken der gesetzlichen Rentenversicherung, in: DRV 1–2, 129.
- Rehfeld, U. G.* (2004): Zur Genese des Forschungsdatenzentrums der gesetzlichen Rentenversicherung, in: Das Forschungsdatenzentrum der gesetzlichen Rentenversicherung (FDZ-RV) im Aufbau, DRV Schriften 55, 12.
- Ruland, F.* (2001): Forschungsnetzwerk Alterssicherung – Verstärkte Kooperation zwischen Rentenversicherung und Wissenschaft im Bereich der Alterssicherung, DRV 3–4.
- Stegmann, M./Luckert, H./Mika, T.* (2005): Die Bereitstellung prozessproduzierter Daten der GRV im Forschungsdatenzentrum der Rentenversicherung (FDZ-RV): Grundsätze zur faktischen Anonymisierung von Mikrodaten und zu Gastwissenschaftler-Arbeitsplätzen, DRV 2–3, 203.