

Working Paper No. 41
ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

Joint ECE/Eurostat work session on statistical data confidentiality
(Luxembourg, 7-9 April 2003)

Topic (iii): Emerging legal/regulatory issues

**DEMAND FOR DATA AND ACCESS TO DATA:
THE RESEARCH DATA CENTRE OF THE STATISTICAL OFFICES OF THE LÄNDER**

Contributed paper

Submitted by the Geschäftsstelle des Forschungsdatenzentrums der Statistischen Landesämter, Germany¹

¹ Prepared by Sylvia Zuehlke (sylvia.zuehlke@lds.nrw.de).

Demand for Data and Access to Data:
The Research Data Centre of the Statistical Offices
of the Länder

Sylvia Zühlke

History

Over the past few years in Germany, there has been a keen debate about access to official microdata for the purposes of scientific research. A commission for the improvement of the information infrastructure set up by the Federal Government came up with suggestions on how to improve the interaction between scientific research and official statistics. Suggestions included the involvement of data users in setting up surveys and processing programmes, plans to update basic and advanced training in the statistical field and various options concerning access of scientific researchers to the microdata produced by public data producers. One of the Commission's central recommendations was to set up a data producers' research data centre as soon as possible. The Statistical Offices of the Länder met this request by setting up a joint Research Data Centre with 16 regional offices. The centre is being run on an associative basis, each office representing a regional site. It is headed by a steering committee, whilst administration and coordination is under the responsibility of the centre's secretariat within the North Rhine Westphalian Office for Data Processing and Statistics. This secretariat constitutes the official contact for the Research Data Centre and is authorised to disclose binding information about the data and services it provides.

The centre's primary purpose

The aim of the centre is to make more data available to researchers and scientists. In Germany, the majority of surveys are carried out on a decentralised basis by the offices of the Länder, which therefore hold the bulk of the statistical data. In order to improve access of researchers, the offices of the Länder will, as a first step, set up an infrastructure for centralising the storage of data. This infrastructure will provide access to official data from every regional site of the Research Data Centre. In addition, an metadata information system will provide comprehensive information about official statistics and the various ways in which they can be used. Furthermore a 'visiting researcher' scheme will be established. In this scheme, researchers may work with anonymised microdata that cannot be transferred as Scientific Use Files for use outside the statistical offices.

Access to microdata

From the scientific research point of view, the ideal solution would be to have as many data sets as possible as 'Scientific Use Files', i.e. to provide researchers with anonymised data sets which can only be de-anonymised by expending a disproportionate amount of time and energy. In the field of household and person surveys such data sets have existed for a number of years now. In contrast, in enterprise surveys there is a completely different situation. Here, the risk of de-anonymisation is much higher than for household and person information data. For this, the Research Data Centre of the Länder will develop an infrastructure which facilitate the use of microdata. It offers three different options of access to data:

- (1) Visiting researchers may use data which are not or only formally anonymised in the context of joint research projects. Here, the basic requirement is that there is a contractually agreed co-operation between one or more statistical offices on the one hand and the researcher on the other hand.

- (2) Visiting researchers can use anonymised data inside the statistical offices. This is the case for data which cannot be provided as Scientific Use Files for the use outside the statistical offices. For such data the input of additional information must be prevented in order to meet the factual anonymity criterion.
- (3) Researchers can use controlled remote data processing: Experts provide the Research Data Centre with analysis programmes for a statistical standard software. The Research Data Centre then applies the programme to the original, non-anonymised, data. Results are transmitted to the researcher after being checked for statistical confidentiality rules.

Demand for data

In order to take into account the scientific demand of data, a special survey was carried out in summer 2002 among potential users of the Research Data Centre. Roughly 600 of the 700 scientists surveyed indicated that they regularly use microdata for their work. The results show the high interest of the researchers in using microdata of official statistics and point to a manifold need of data. Concerning the different options for using data there is a high preference of the use of Scientific Use Files. Yet, as Scientific Use Files always led to compromises concerning the information content and as their production takes a lot of time and energy, this option is not suitable for every research project. Thus, the survey has shown potential conflicts between the different types of demand for data and the preferences for the different options for using data. A solution to this problem is offered by the new decentralised infrastructure for the visiting researcher scheme and a combination of the different options of data use.

In addition to the work relating directly to improve the access to data, the centre will also be addressing basic questions of data access. These include, for instance, the question of statistical confidentiality in the face of an ever-increasing volume of readily available additional information, and the expansion of cooperative links with other national and international research data centres.

Contact

Dr. Sylvia Zühlke
Geschäftsstelle des Forschungsdatenzentrums der Statistischen Landesämter
c/o Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen
Postfach 10 11 05
40002 Düsseldorf
Germany
Fon: 02 11/9449-2547
sylvia.zuehlke@lds.nrw.de