

Documentation

Towards an Improved Statistical Infrastructure

**Summary Report of the Commission
set up by the Federal Ministry of Education and Research (Germany)
to improve the statistical infrastructure in cooperation with
the scientific community and official statistics***

Major results of stocktaking

Assignment of the Commission

It is only on the basis of valid and reliable data that social scientists, including economists, can carry out realistic studies and develop convincing recommendations for policy action. In this context, data availability and quality are a central precondition. Hence, the performance of the data infrastructure is a decisive basis for both the performance of a society and the innovation capability of social and economic research on an international scale.

Statistical surveys cause costs and impose a burden on citizens and businesses. For this reason, the existing data, including those which have been collected for other administrative purposes, should be used as effectively as possible.

The way statistical data are used has changed in the information society. Changes in the economy and society can only be studied if information is available at the microlevel of observation units (persons, households, businesses) and accessible for statistical purposes at varying times. Modern sta-

* Members of the Commission were Johann Hahlen, Hans-Juergen Krupp, Gerhard Arminger, Richard Hauser, Eckhart Hohmann, Karl Ulrich Mayer, Walter Mueller, Axel Reimann, Willem F.M. de Vries and Gert G. Wagner. For the full report, written in German, see „Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (Hrsg.): Wege zu einer besseren informationellen Infrastruktur. Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Baden-Baden: Nomos Verlagsgesellschaft 2001“ (ISBN 3-7890-7388-1). The appendix of this report contains an English language summary which is reproduced here (recommendation 36 which deals with specific funding questions within Germany is not included).

tistical methods require access to microdata to analyze behavioral routines and relations.

In accordance with her responsibilities in the area of social science and scientific infrastructure, the Federal Minister of Education and Research asked a Commission to examine whether the informational infrastructure in the Federal Republic of Germany is still adequate to fulfil its functions that have changed as mentioned above. She entrusted the Commission with preparing proposals for improving the informational infrastructure between the scientific community and official statistics. The assignment of the Commission is aimed at generally improving cooperation between the statistical and scientific communities. It refers to all data sources and above all is to rely on international experience and, in this context, to check whether examples of good practice may be applied in the prevailing situation.

Society's demand for scientific-statistical information

An international comparison of best statistics and best procedures of statistical analysis shows that the informational infrastructure in the Federal Republic of Germany has shortcomings that could be improved on:

- There are considerable gaps of information in the sectors of the economy and society which are subject to rapid change.
- It is only in few areas that official statistical authorities collect data repeatedly from the same statistical units. This has led to a deficit in both longitudinal data and dynamic analyses based on them.
- Science-based longitudinal studies have a fragile financial and institutional basis while, in other countries, they are financed by public funds in the long run.
- The scientific and statistical analysis of register data is often not sufficiently detailed, and the data are largely not accessible to social scientists. Linking the data of different registers for statistical purposes is either hardly or not at all possible.
- The consideration of many relevant issues requires access to microdata which, however, is not sufficiently ensured for numerous stocks of data.

Fulfilling the data requirements in various fields of information

In the context of stocktaking, the Commission has dealt with the fulfillment of data requirements in different fields of information. For this purpose, it has collected expert opinions on the situation in the different fields

of information and subsequently analyzed the observations included and assessments made of data deficits and requirements. The following *fields of information* have been covered: population; government and politics; employment; income and property; healthcare and social welfare system; consumption and housing; education and culture; economy and transport; environment and energy; science and research; subjective indicators.

Reports submitted by experts show that the information offered is very detailed and diverse. Nevertheless, on the one hand, its structural arrangement is inadequate so that even experts face difficulties in keeping track of information in their own areas. On the other hand, there are considerable information deficits.

Overview of the informational infrastructure in Germany

The commission's report gives an overview of the major providers of statistical information in Germany. Following a brief description of the information providers, the stocks of data they hold and their practice of giving access to anonymized microdata are described.

The publicly funded statistical infrastructure goes far beyond official statistical data as produced by the Federal Statistical Office, the statistical offices of the *Länder*, and municipal statistical offices. Statistical data are also collected by ministries at the federal and *Länder* level in the context of ministerial research and by other government bodies and paragonmental institutions such as social insurance funds. Besides, stocks of statistical data are compiled in the process of administrative procedures. Institutions for social science research which are financed by public funds (e. g. universities, institutes of Max Planck Society [MPI], institutes of the Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz [WGL] and other publicly funded research institutes), too, conduct surveys of various types whose results enlarge the overall stock of existing empirical data. And finally, private market and social research institutes and also commercial providers of information are included here.

Cooperation and communication of official statistics with the scientific community

The conglomerate of different producers and stocks of statistical data has a historical background. So far, neither a co-ordinated specification of the data collection and processing programs nor a general accessibility of the data stocks or a systematic storage in archives have been ensured.

Though cooperation and communication between the scientific community and official statistics have been partly institutionalized on the part of official statistics, particularly by the Statistical Advisory Committee at the Federal Statistical Office (*Statistischer Beirat*), this has not been the case on the part of the scientific community, which has failed to develop an adequate legitimization procedure for an institutional cooperation with official statistics. The scientific community has nearly no regulation-based opportunities to influence ministerial surveys or the specification of process-produced data. Though having other priorities of work, the way in which public bodies, which are actually part of the statistical infrastructure see themselves with respect to their statistical activities, is often not clear. Though data are stored in archives in many places, such storage has gaps and is not based on an integrated concept.

The advantages of a coordinated and consistent concept of the publicly funded collection, provision and archiving of data are the bigger and uncoordinated double surveys are the better avoidable, the more comprehensive, cheaper and faster the access to existing data stocks is for scientists outside the data-producing institutions (in accordance with the rules of privacy protection).

The Commission appreciates that cooperation between official statistics and the scientific community in universities and research institutes has considerably improved in the past few years. This development has been reflected by manifold activities typical examples of which are the following: the publication of a "Social Monitoring Report" (*Datenreport*) which – since 1983 – has been jointly compiled by the Federal Statistical Office (*Statistisches Bundesamt*), the Science Center Berlin of Social Research (*WZB*) and the Mannheim Centre for Survey Research and Methodology (*ZUMA*); the activities of the microdata department at *ZUMA*; setting up a new forum for discussion with representatives of the major German economic research institutes; organizing colloquiums and workshops together with the German Statistical Society (*DStG*), the Working Party of Social Science Institutes (*ASI*) and the Working Group of German Market and Social Research Institutes (*ADM*), and with individual scientists.

Ways of transmitting data to scientists not working within the system of official statistics have been tested by providing de facto anonymized microdata files of the Current Population Survey (*Mikrozensus – MZ*), the Income and Expenditure Survey (*EVS*), the employment survey (*Beschäftigtenstichprobe*), the time use survey, and the European Community Household Panel. Internships and research visits at the Federal Statistical Office and the Institute for Employment Research at the Federal Labour Office (*IAB*) (partly in the framework of pilot projects of the Federal Ministry of Educa-

tion and Research [*BMBF*] or promoted by the German Research Community [*DFG*]) are further welcomed measures to improve the relations between official statistics and the scientific community.

Specifying data collection and processing programs

Unlike in most other industrial countries, the *decision* on the type of official statistical data to be collected in Germany is not taken by the government or statistical offices, but the legislators. This situation does not only make it more difficult for the system of official statistics to adapt its statistics to changing conditions, but has also hindered an inclusion of an optimum degree of scientific expertise into the data collection programs.

Archiving

A problem which is of particular importance to scientists is the long-term storage of data in archives. In many cases, historical statistical data are no longer accessible (for either physical or software-related reasons). The results of archiving the data of the first Income and Expenditure Survey (*EVS*) in the framework of a university research project (Johann Wolfgang Goethe University at Frankfurt am Main) show that a successful long-term storage has so far depended on (good) luck.

Educational deficits at and outside of universities

There is concern regarding a variety of deficits in higher education which result in either an insufficient or an improper use of the existing infrastructure. Workable empirical analyses require a linkage between substantial scientific expertise, a clear understanding of the collection and processing of data, and a problem-oriented statistical or econometric analysis. The basic knowledge students have of descriptive statistics, simple statistical modeling, simple inferential statistics and econometrics, and the basics of applied computer science, in particular database systems is inadequate in many institutions of higher education. Apart from that, there often is a lack of knowledge of software packages for statistical analyses.

Many scholars teaching economics, business administration, social sciences, but also mathematical statistics and econometrics are not involved themselves in empirical matters so that they are not in a position to supervise activities of that kind. University faculties of economics, business administration and social sciences have seen a decline in chairs (full professor-

ships) with regard to statistics, econometrics and empirical social and economic research or a revaluation of chairs down to the level of associate professorships, and the number of research and teaching assistants has generally declined. This has contributed to an aggravation of the above problems.

It is particularly in economics that issues like collecting, processing, coding and transforming data are not taught systematically. Existing data sources such as the data provided by the Federal Statistical Office or by the Central Archives for Empirical Social Research (ZA) in Cologne are not used to a sufficient extent. Besides, the availability of public use microdata files to train students in using this kind of data is not satisfactory.

Scientists' access to (micro)data

Legal frame in Germany

Instruments regulating data protection are e.g. the Convention of the Council of Europe at the *international* level, the Data Protection Directive of the European Union at the *Community* level, the Federal Data Protection Act (BDSG) at the *national* level, and data protection acts at the *Länder* level.

Two of the major principles relating to the framework conditions for science and research are the *principles of consent* and of *earmarking*.

The *principle of consent* means that individual data must not be processed without the prior approval of the individuals to whom they relate, unless processing has been initiated by and is based on a rule of law.

The *principle of earmarking* requires that individual data must be only processed in the framework of previously clearly defined (and legally permitted) purposes.

Those regulatory principles may collide with individual research designs or general methods of research. However, there are exceptions to each of the principles. The specific interests of the scientific community are served by the following *three approaches*:

- modifying the reference to individuals,
- modifying the principle of consent, and
- modifying the principle of earmarking.

The *reference to individuals* may be eliminated by anonymization. Due to additional knowledge, however, a complete anonymization can hardly be achieved. If, however, a risk assessment yields a small risk of reidentification, the legal concept of “reference to individuals” will cease to apply pro-

vided additional measures are taken to cover that risk. At the same time, the problem of reference to individuals may be eliminated by depositing data in the office of a notary (*Treuhänderlösung*).

Consent is not required if the processing of data is permitted (among other things) by a special legal provision.

Some laws of data protection see scientific research as autonomously legitimated. Such approaches require, as a compensation, that the interests warranting protection of those concerned should be thoroughly weighed up.

In the Federal Republic of Germany, the specific data protection regulations aimed at statistical confidentiality are both in line with European legislation and largely determined by the “population census judgement” of the German Federal Constitutional Court. They stipulate the following:

- a) The possibilities of data transmission shall be laid down by law.
- b) Microdata shall be de facto anonymized.
- c) The recipients of the data shall be incumbents, persons specially sworn in for public service, or persons sworn in in accordance with the Law on the Commitment of Persons to Secrecy.
- d) The addressees of transmission in institutions of higher education or other institutions shall be engaged in independent scientific research.
- e) Microdata shall be deleted as soon as the scientific project has been completed (Art. 16 (8) p. 2 of the Federal Statistics Law [*BStatG*]).

The provisions of Art. 16 (6 and 7) of the *BStatG* of 22 January 1987 correspond with the above principles.

The following strategies have been developed for the provision of microdata as part of the cooperation between statistics and the scientific community:

1. development of de facto anonymized Scientific Use Files,
2. methodological integration instead of data transmission
 - a) by incorporating scientists into the institutions of official statistics,
 - b) by “remote computation” based on software of the researcher.

The generally feasible solution of transmitting data based on consent has not played an essential role so far.

Besides, the specific data protection regulations to ensure confidentiality of statistical data cover any micro-level data, not only data relating to individuals. Above all, this also refers to business-related microdata.

The principle of interpretation of the Constitution’s unity requires regulations which include a normative concretizing of weighing the trade-off be-

tween the individual right of data protection and data access which is a right of the scientific community. The existence of this Commission itself proves that, despite considerable progress, the existing regulations are not sufficient. This is mainly due to the French model of regulation of the general data protection laws of the Federal Republic of Germany which has been designed as a prohibition with reservation of permission (*Erlaubnisvorbehalt*).

There has been a legal discussion on granting the scientific community an extended data access based on the creation of “research data confidentiality” (*Forschungsdatengeheimnis*). However, this would require the following in addition to the stipulations of Art. 16 (6, 7 and 10) of the BStatG:

- an institutional segregation,
- seizure-resistant data processing,
- a prohibition to issue and transmit data outside the given sector,
- a clear definition of the research area, and
- other measures ensuring basic rights such as the right of those concerned to information, codes of conduct, codes of ethics, and independent checks of data processing.

International exchange of data and the Safe Harbor principle

It is a principle of data exchange at the international level that such exchange between countries does not have to meet additional data protection requirements only because the data cross the borders between countries, provided the data protection regulations in the countries concerned operate to a similar effect. As regards supplies by Member States of the European Union (EU), these are subject to the stipulation that individual data must only be provided to non-EU members who will ensure an adequate protection of privacy.

After the European Commission had concluded its negotiations with the US trade department, involving the European Parliament and lasting over a period of more than two years, the European Commission took a decision on 26 July 2000 regarding the Safe Harbor principle. According to that decision, the Safe Harbor concept as followed by the US trade department ensures an adequate protection of the individual business data that are transmitted from the EU to the USA.

The Safe Harbor concept which has been in operation since 1 November may serve as a model for the exchange of statistical micro data in the Member States with institutions in third countries such as the USA., Canada, Australia, Japan, etc.

*Organizational and technical facilities and de facto access
to aggregated data in Germany*

In accordance with their legal assignment to publish statistical information and the way they understand their responsibility to disseminate their data in a tailor-made manner, the statistical offices at the federal and *Länder* levels offer aggregated data of all federal statistics in manifold ways. In this context, the following main forms of data provision may be differentiated:

1. basic information offered free of charge;
2. statistical products and services;
3. tailor-made solutions for individual users.

The last of the three forms of information provision refers to problem-oriented and tailor-made products and services which are specially produced on the request of individual clients and therefore are subject to payment of all costs incurred.

The typical users of statistical data in the past were final consumers who needed the data for their own purposes. The data offered have however become increasingly attractive to commercial providers of information, too. Upon conclusion of a license agreement including the required payment stipulations, statistical data are therefore also released to third parties for being marketed by them.

*Organizational and technical facilities and de facto access
to microdata in Germany*

Transmitting records of data collected at the microlevel of observation (individuals, households, businesses) is not part of the traditional tasks of official statistics in Germany. However, the legislators have in this respect granted the scientific community an exclusive privilege among the users of official statistics.

In line with that privilege, the scientific community may receive microdata records from the stock of official statistics. There are four basic opportunities which are offered for their use:

- (1) Scientists receive “de facto anonymized” microdata

Since the latest amendment of the Federal Statistics Law (*BStG*) in 1987, the scientific community, unlike other users of official statistics, has had the

opportunity to use data in a de facto anonymized form in addition to “absolutely anonymized” individual data (*Wissenschaftsprivileg*). Since that time, access has been provided to de facto anonymized microdata in many cases. However, process-produced administrative data often are subject to the specific provisions and restrictions of social data protection.

De facto anonymized microdata are provided by the statistical offices based on specialized analyses of the statistical material in accordance with individual and specific data requests. In some cases where many users request similar statistics, the statistical offices supply de facto anonymized microdata files in the form of standard products (scientific use files). In both cases scientists receive anonymized microdata which they may use for their research purposes without any restriction.

De facto anonymized primary files, which include data of all federal *Länder*, have been produced by the Federal Statistical Office in cooperation with the statistical offices of the *Länder* for the following household and individual-related statistics:

- *Mikrozensus* (Current Population Survey); 1989, 1991, 1993, 1995, 1996, 1997 (1998 planned for April 2001);
- *EVS* (Income and Expenditure Survey); 1993, 1998;
- German sample of the European Community Household Panel; 1994 until 1996;
- time use survey; 1991/92.

Since the anonymization of microdata is generally rather expensive, such data are made available exclusively against payment as stipulated in the respective cost regulations of the Federal Statistical Office. Scientists in Germany may obtain de facto anonymized individual and household data from the Federal Statistical Office against minor fees based on a pilot project which was carried out together with the Federal Ministry of Education and Research and the Society of Infrastructure Institutions of Sociology and Political Science (*GESIS*). The Federal Labour Office, too, provides access to de facto anonymized microdata pursuant to Art. 75 of Social Code X (*SGB X*). In this context, the 1% sample of employment statistics should be especially mentioned whose results are made available in the form of anonymized data (*Beschäftigtenstichprobe*) to scientists in cooperation with the Central Archives for Empirical Social Research (*ZA*). Currently, further-going supplies of anonymized data from primary statistics of the Federal Labour Office are only feasible upon checking each individual request, against payment, and to the extent permitted by the resources available. The German Bundesbank which, in addition to similar banking secrecy provisions applies the regulations of the Federal Statistics Law with respect to

releasing microdata, has not yet transmitted any of its micro data for scientific research projects.

(2) Scientists provide their own programs for analyzing purposes

This form of cooperation is characterized by the fact that scientists themselves develop analyzing programs for their research purposes, which they transmit (e.g. via teleprocessing) to the competent office storing the microdata required. The office makes available the internal microdata requested in the necessary form and applies to them the programs supplied by the scientists. In this way, scientists do not have any contact with the microdata to be kept secret; the microdata are not subjected to any anonymization procedure.

(3) Scientists analyze non-anonymized microdata

Another form of using the stocks of official statistical microdata is research based on *original statistical data*. This form, however, may give the persons involved the chance to identify rather easily the identity of the data provider. In accordance with the current legal stipulations, the *statistical offices* are therefore of the opinion that using data in the above manner may be permitted only if the statistical office concerned initiates the research project itself or has a very strong interest in the issue to be investigated by an external scientist so that it will use the results of the project for its own purposes.

The form of cooperation described above is based on the relevant stipulations of the Federal Statistics Law. However, the interpretation of these provisions by the *statistical offices* differs. The majority of them holds the view that scientists, who have been specially sworn in to observe statistical secrecy, could be permitted to work with non-anonymized microdata in statistical offices *without* being commissioned by a given office to do so only if the relevant legal provisions were modified. Thus without such modification of the Federal Statistics Law, a uniform approach to this problem, which will better serve the interests of the scientific community, cannot be achieved.

(4) Specialized analyses of statistical material commissioned
by the scientific community

Specialized analyses of statistical material by statistical offices are certainly not a form of data exploitation that is preferred by external scientists.

However, this method should be mentioned here as well as it may serve to exploit the informative potential of microdata in order to answer scientific questions where due to data protection other methods cannot be applied.

Access to microdata in non-commercial research

The replicative data collection programs of social sciences and economics follow the principle that the data-collecting institutions and researchers make the data available either immediately or after initial analyses to the scientific community free of charge or for a small nominal fee. In Germany, with the exception of the Socio-Economic Panel (SOEP), the data producers usually transmit their data to the Central Archives for Empirical Social Research (ZA). The ZA stores the data in archives and disseminates them to scientists and other users. The SOEP data, however, are transmitted to users directly by the producer. On the whole, empirical research in the area of social sciences is largely characterized by the principle that the data collected should be made available in an adequate form to other scientists, though partly after a period of protection for the producer's own analyses. In many cases, this principle of re-analysis results in a very intensive exploitation of data records. It creates the basic precondition for subjecting results of individual researchers to a data-based criticism by other scientists, developing and checking alternative models for explaining reality and, above all, depicting social and economic changes over time. In this context, service functions like those performed by the ZA are indispensable.

As a matter of fact, there are still cases in which scientists are not prepared to make available the data collected to the scientific community as a whole. For this reason scientists, and especially researchers whose sample surveys are financed through public funds – in particular the German Science Foundation (*DFG*), should as a rule be obliged to make available the microdata obtained in an adequate form to other scientists for purposes of re-analysis.

Using commercial data

In the past two decades, a considerable range of statistical data has been offered by commercial institutions which, as a result of the structure of demand, encompass mainly economic information. Management consultants, market research institutes and businesses in the information industry supply data bases which they have produced themselves. The sources of those data bases are, on the one hand, official statistics and, on the other, surveys which those institutions conduct themselves.

Commercial providers process the data of the Federal Statistical Office, the Bundesbank and other public institutions, which act also as providers of their own databases – where necessary – through homogenization (adjustment of data based on different definitions) and aggregation. For commercial producers and providers, the market and their clients' demand are the basis for orientation.

The importance of such databases for the scientific community has so far been rather small, although it is expected to grow.

Actual situation of data linkage in Germany

The situation of statistical data linkage is not satisfactory in Germany. Although data linkage is not a problem in terms of data protection if every single respondent, or unit covered, gives his consent to such linkage, the ensuing procedure is largely unfeasible or at least inefficient. The reason is that, if approval is explicitly required, many persons or units concerned will not give their consent even though the linked data cannot be allocated to individuals / units.

International comparison regarding access of the scientific community to microdata of official statistics

The Commission performed a large-scale international comparison showing the current situation of access to microdata in France, the United Kingdom, Canada, the Netherlands, Norway, Austria, the U.S.A., Denmark and Sweden. The following issues are discussed for any of the countries examined:

- Microdata of official statistics available for research;
- Access to microdata by the scientific community;
- Procedure of making data accessible, conditions for use, and pricing;
- User support;
- Cooperation between official statistics and the research community.

Altogether, presenting the situation in the selected countries provides a highly varied picture of cooperation between the scientific community and statistical offices. The conditions of informal and institutional cooperation vary considerably between the countries. In a number of countries, the closeness between statistics and the scientific community and, consequently, the cooperation between the two spheres is much stronger than in Germany. The examples from the examined countries also illustrate where and how such cooperation can be improved – even with different institutional conditions –, so that both sides and the society as a whole will benefit from it.

Recommendations of the Commission**

Improved cooperation between science and statistics is necessary for data users and data producers

1. *The Commission recommends that the division of labor between official statistics and science-based data production – which generally has proved suitable – should be maintained, whereas the coordination between the scientific community and official statistics should be improved through institutional regulations.*

Over the last twenty years, social and economic sciences in the Federal Republic of Germany, in cooperation with commercial market and public opinion research institutes, have set up a highly detailed microdata infrastructure based on representative sample surveys. Reasons for that development have been the difficult access to microdata of official statistics and, in particular, the specific opportunities provided by data collected through science-based surveys.

Therefore, it must be a major goal for scientific research to combine the advantages of data from survey research (theory orientation, sensitivity to the system of measurement, topicality of issues examined and timeliness of data access, multivariate analysis of microdata, longitudinal studies) with those of data from official statistics (complete enumerations or large-scale samples, cohort series, regional breakdown, small groups, reliable historical comparisons, contextual relation to households, residential districts, place of residence and place of work, supply of data which are necessary for re-adjusting sample surveys).

As public funds must be used as efficiently as possible, it is necessary that official and science-based statistics be coordinated by scientific and public institutions. However part of such coordination might become unnecessary if all statistical microdata that are financed by public funds were made available without restrictions (while ensuring data protection) to the entire scientific community for re-analyses.

2. *The Commission recommends setting up a “Council for Social and Economic Data” in order to assess and improve the “informational infrastructure”*

The Commission recommends that a Council for Social and Economic Data be set up as a federal controlling and initiating body in which major data producers and data users as well as the Federation and the *Länder*, in

** This abridged version does not contain each recommendation in detail. The Commission explicitly refers to the detailed chapter E of the original report (written in German – see footnote * above) dealing with recommendations.

their function as providers of funds and locations, would have to be represented.

The functions of such a Council for Social and Economic Data would be to

- assess and improve the data infrastructure in Germany and to develop recommendations for the Federation and *Länder* on the program of science-based statistics and its funding,
- promote social and economic reporting and permanent monitoring of societal issues,
- give advice and recommendations for, and evaluate the setting up and the activities of research data and service centers (recommendations 29 and 33),
- allocate project funds, or suggest them for allocation.

The plenum of the Council for Social and Economic Data should establish a working committee to be supported by professional staff.

Participation of the scientific community in developing survey and processing programs

3. *In order to organize systematic participation of the scientific community in developing survey and processing programs of official statistics and in the specific hearings held at the parliaments (including data protection legislation), the Commission recommends that the Council for Social and Economic Data hold discussions in cooperation with the relevant scientific associations (in particular Deutsche Gesellschaft für Soziologie, Verein für Socialpolitik, Deutsche Gesellschaft für Bevölkerungswissenschaft, Deutsche Statistische Gesellschaft, Deutsche Vereinigung für Politische Wissenschaft, Schmalenbach Gesellschaft, Deutsche Forschungsgemeinschaft [DFG], Max-Planck-Gesellschaft [MPG], Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz [WGL]).*

Regarding the definition of survey and processing programs of official statistics, justified complaints have been voiced that scientific considerations have not enough influence there. However, this is not only due to the fact that the system of official statistics is not sufficiently open to new ideas. There are two other reasons: First, the survey and processing programs of official statistics are in part inflexible only because they are too much subject to legal details. Second, the discussion within the scientific community on the development of survey and processing programs of official statistics and on an optimal supply of science-based statistical data is in need of improvement.

4. *The Commission recommends that the legislator limit legal provisions in the laws on certain statistics to the minimum required by the constitution and leave it to the statistical offices and their advisory bodies to define details of survey programs.*

The statistical offices can take account of new issues and ideas from the scientific community only to some extent, which is determined by how much room for manoeuvre they actually have in shaping the survey programs.

5. *The Commission recommends that the legislator extend the definition of the Statistical Advisory Committee's tasks as laid down in the Federal Statistics Law to include medium-term program planning (Statistischer Beirat). It further recommends that the legislator provide for compulsory hearings to be held before introducing, abolishing or substantially changing specific official statistics. The number of representatives of the scientific community should be increased and it should be ensured that representatives of empirical social and economic research are sent to the Advisory Committee.*

Such a regulation would have to ensure that the scientific community is represented reasonably and with sufficient expertise in the Statistical Advisory Committee, that the committee is informed in time before laws are adopted regarding planned modifications to programs, and that it is in a position to voice and defend its interests also in the course of the legislative process concerning data protection.

6. *The Commission recommends that the legislator create regulation-based opportunities for scientific advice with regard to surveys conducted by ministries and non-statistical authorities, agencies and institutions such as Deutsche Bundesbank, Bundesanstalt für Arbeit and social security institutions.*

Priorities in continuing and developing important statistics

The Commission considered it one of its tasks not only to make procedural proposals but also to give subject-related comments on maintaining and developing the informational infrastructure. The following reforms appear particularly urgent to the Commission.

In the field of *official statistics*:

7. *The Commission recommends conducting a population census in order to obtain not only the basic data that are of immediate interest but also a basis for reliable breakdowns of the population and expansion of all sample surveys.*

8. *The Commission recommends that the Mikrozensus (Current Population Survey of Germany) be further developed. This regards switching over to an infra-annual survey concept which should allow analyzing the data as a rotating panel. Usefulness would considerably be improved if exact data on gross earned income were collected. On that basis, an access panel could be developed for drawing voluntary household samples. However, that activity will yield its full benefit only if, based on the success of the pilot projects, the scientific use files of the Mikrozensus are continued.*

9. *The Commission recommends improving the sample survey of income and expenditure (EVS) – in methodological terms by reducing the time intervals and introducing a rotating sample concept and, in subject-related terms by presenting detailed wealth data in order to obtain information on private-sector coverage of social risks.*

10. *The Commission recommends bridging serious gaps in business sector statistics.*

What is particularly urgent here is the further development of statistics about the service economy. Also, business modifications such as start-ups and concentration processes require better statistical coverage.

In the field of *science-based statistics*:

11. *The Commission recommends that the German Socio-Economic Panel Study (GSOEP) be institutionalized on a permanent basis as a wide-range, science-based and sufficiently large longitudinal sample survey.*

Because of its great information value both for scientific research and policy analysis and because of its widespread use, the SOEP is indispensable for the informational infrastructure in Germany. Implementing the recommendation of the High Scientific Council (*Wissenschaftsrat*) regarding permanent institutionalization and funding is urgent. What is also necessary is extending the sample as this would make it much better suited for research on small sub-groups. In this context, it will be important that the SOEP can play a role in the social reporting system envisaged by the EU for the European level.

12. *The Commission recommends continuing ALLBUS (German General Population Survey), ISSP (International Social Survey Program) and welfare surveys (Wohlfahrtssurveys), as they are important data sources for the social sciences.*

ALLBUS and ISSP data are the survey data that are frequently used in social sciences and are of substantial importance for higher education. The

welfare surveys have become an important tool of permanent social monitoring and should be continued in a suitable manner.

13. The Commission recommends stronger support of cohort studies, such as longitudinal studies of human development.

Providing cohort and person-related longitudinal data, prospective and retrospective cohort studies are especially well suited to cover and analyze processes of human development and life in the context of social and economic change. As highly detailed individual data are thus available, special data organization activities are required to make the results of such studies available to more general scientific research. What should be supported is both continuing existing cohort studies and generating new cohort studies that would cover early childhood, adolescence and early adulthood.

The Commission would like to restrict its recommendations to this small number of particularly urgent projects. As shown in the full report (written in German), there is a large number of other areas where improving the informational infrastructure would be useful. Relevant decisions should be made following the procedures suggested by the Commission.

Supporting research on data collection, processing and archiving

14. The Commission recommends that the Federation request the setting up of a special commission of the High Scientific Council (Wissenschaftsrat) on the current state of affairs in higher education and research regarding the methods of empirical social and economic research or that the Federation explicitly define that area to be a task of the commission currently examining economics (high commission on “empirical economic research”).

As is shown by international comparisons, Germany has deficits in higher education and research regarding the methods of empirical social and economic research. The Commission recommends setting up chairs or centers at institutions of higher education which – both in teaching and research – would especially be oriented towards methodological problems of survey statistics and official statistics.

15. The Commission recommends – for the purpose of supporting methodological research in official statistics – that the science community shall be more involved in the further development of methodological instruments and that the budgetary conditions be created to allow sustained methodological research in official statistics.

The statistical offices are not primarily administrative authorities but most of all institutions which perform scientific tasks. This component of

their work is reflected neither by the tasks assigned explicitly to them nor by their budgets. In order to make use of scientific expertise for official statistics and to provide new ideas for methodological improvement, joint research projects by scientific and official statistical institutions should be carried out much more often than in the past.

16. The Commission recommends that the Council for Social and Economic Data that is to be set up should be commissioned to deal with the problem of archiving statistical data.

17. The Commission recommends that the DFG establish the subject of "Empirical Economic Research" (similar to "Empirical Social Research" as a sub-discipline [or as an extension of the "Statistics" sub-discipline]).

This means that reviewers for "Empirical Economic Research" would explicitly be elected. In normal DFG procedures and beyond, this would contribute to achieving a reasonable representation of empirical economic research in the self-governing bodies of (German) scientific institutions and to enhance and ensure over the medium term the development of the subject at universities.

Higher education and training

18. The Commission recommends that the universities and faculties work towards improving the education of undergraduates in the areas of statistics, econometrics and applied computer science by using realistic data sets. In graduate studies, intensified instruction is required on statistics and econometrics as well as on data collection, processing, protection and evaluation. Establishing regular graduate programs according to the Anglo-Saxon model is of crucial importance for scientific top performance. In this context, learning new empirical methods and in-depth studying of statistics and econometrics should be compulsory for economists and social scientists.

The involvement of official statistics in higher education should be enhanced by granting more teaching assignments to improve knowledge of specific methods and databases of official statistics.

19. The Commission recommends that universities and ministries of science cancel the measures aimed at reducing the number of professorships in empirical social and economic research, statistics and econometrics in the relevant faculties and upgrade a large number of existing associate professorships to full professorships in order to make working in those areas more attractive.

20. The Commission recommends that seminars and advanced training courses be organized in cooperation between universities and statistical agencies. This should even include further education and advanced training programs or summary schools for young scientists and junior staff in official statistics. All this might be part of the specific aims of non-university institutes.

For example, GESIS (German Social Science Infrastructure Service) should explicitly be commissioned to extend its range of advanced training courses offered, e.g. through interdisciplinary summer schools in cooperation with suited university teachers, offering instruction in sampling methods, collection methods and data analysis for the spheres of official and non-official microdata.

Economic aspects of data access

21. The Commission recommends that aggregated data of official statistics be made available via Internet, which – at least for the scientific community – would be largely free or subject only to a small nominal fee.

Additional requests by internet hits involve practically no additional costs. It is a normal feature of research activity that at first one does not know what data are suited for a specific purpose. It is only in the course of the research process that it emerges what data are really required. If, however, every single request is subject to charge, considerable amounts will add up rapidly.

22. The Commission recommends, following the example of, and continuing the pilot projects supported by the Federal Ministry of Education and Research, to reimburse to the data producers the fixed costs of anonymization (including documentation, i.e. of producing Public Use Files for the scientific community and of Scientific Use Files) at a flat rate from research funding. The individual researchers would then have to pay the marginal costs of actual data delivery.

As apportioning the supply costs for Scientific and Public Use Files to the individual customers would have a prohibitive effect on the large majority of scientists, the fixed costs are covered by research funding jointly for all users. Such a pricing model is beneficial also to the data producers because it ensures that the fixed costs of producing such files will be reimbursed to them.

Access to aggregated data

23. *The Commission recommends that a joint database system of official statistics be set up that may be accessed via internet and that contains the data of all federal statistics broken down by regions.*

As regards the access to the databases of the statistical offices, there is considerable need for modernization. It is not very user-friendly that the various statistical offices within Germany all have their own databases with different user interfaces and that material in a detailed regional breakdown is not available at one central agency.

Microdata access and data protection

24. *The Commission recommends that different ways of accessing microdata should be used, depending on the kind of data.*

Microdata are data on individual persons, companies and organizations. They provide the data while trusting in anonymity and confidentiality; therefore effective measures are required to protect such data.

An inherent element of scientific methodology is the principle that, once submitted, results must be able to be verified. Consistently observing that principle is possible only if the methods originally applied to obtain the results are available for re-analysis.

25. *The Commission recommends that procedures be developed which allow checking whether specific users are authorized for use of data and whether they fulfill the preconditions. What is desirable is the development of a code of conduct permitting sanctions.*

Technical and other provisions on the part of the data users serve the purpose of data protection. Those measures ensure that the data are not accessed by unauthorized parties and that no attempts can successfully be made to reidentify those who provided the data. Necessary is, for instance, a periodical revision of the list of technical measures developed as part of the German Anonymization Project (carried out at the University of Mannheim). A code of ethics, as practiced in many disciplines now, should describe in detail the obligations of scientists and research institutions regarding adequate behavior in terms of data protection when using microdata. Such a code of ethics should be developed in cooperation of the disciplines concerned.

What should also be clarified is which users would benefit from the privilege granted by the data protection law to scientists. For institutions whose status cannot clearly be defined, the tool of certification has to be created.

26. *The Commission recommends that the development of Scientific Use Microdata Files be enhanced as they are the most important tool of microdata access. To allow performing analyses of social change, it is recommended that – for those surveys where Scientific Use Files have now been created – suitable older data be made available in that way, too, and that similar files be provided such as regional Mikrozensus files or Mikrozensus panel files.*

De facto anonymized microdata which are made available to scientific institutions in a standardized manner for research purposes are without any doubt the most cost efficient way of data access. At the same time, it generally is the most suitable and most flexible solution for researchers. So far, Scientific Use Files have been created mainly for personal and household data. For data on organizations and enterprises, however, there is still considerable need for research in terms of anonymization strategies.

27. *The Commission recommends carrying through a joint research project of the scientific and the official statistical communities on possibilities of de facto anonymization of data on enterprises and local units.*

The purpose of that project would be to study, in the light of the more recent methodological and technical measures, whether the opinion that economic-statistical microdata in general are not suited for anonymization is still justified – an opinion frequently voiced in a very much generalized manner – or whether, in spite of the well-known problems, anonymization strategies can be developed to permit access to enterprise and local unit data, too, to the scientific community.

28. *The Commission recommends the development of Public Use Microdata Files.*

What distinguishes Public Use Files from Scientific Use Files is mainly their more thorough, “absolute” anonymization. Hence, passing on such Public Use Files is not linked to the privilege granted to scientists. It could be beneficial to use Public Use Files for student training; they could cover a major part of commercial user interest and they could help solve at least some of the problems involved in the present exclusion of foreign scientists from microdata use.

29. *The Commission recommends setting up Research Data Centers (RDCs) with the opportunity of controlled teleprocessing*

The crucial limit to the usefulness of Scientific Use Files is data files for which de facto anonymization alone would substantially impair the information value of the data. The same is true of studies that require matching of various data sets. For such data and activities, it is recommended that Research Data Centers (RDCs) with special safety measures be set up, which

have proved quite successful in the USA. Scientists can do their studies there. Setting up such RDCs seems to be the only feasible way to make available for research purposes particularly sensitive data that can hardly be anonymized, especially at the local unit or enterprise level.

Generally, RDCs offer data of various data owners and data producers. In particular, they are not restricted to data from official statistics. They should also make available data from scientific research that require particular protection and cannot sufficiently be anonymized. RDCs should be set up with the data holders concerned and also, subject to adequate control by the data holders, at individual scientific institutions. However, this must not lead to decentralizing research data centers by data holders.

At RDCs, it should be possible to do controlled teleprocessing. Scientists may transmit their requests for data analysis by electronic means to a RDC. Having checked the requests in terms of data protection, the staff of the RDC will perform the requested analysis and transmit the analysis result – again checked in terms of data protection – to the requesting scientist.

30. The Commission recommends that the data producers implement a visiting scientist model.

It is quite obvious that setting up RDCs will take a considerable amount of time before they are fully functional. It is meaningful to set up facilities already now where researchers – with commitments regarding data protection – have the same status as staff members of statistical offices or other institutions holding data and may perform microdata evaluations while staying there as visiting scientists. As regards the selection of guests, the Commission recommends to develop a transparent procedure.

Using international microdata

31. The Commission recommends improving the situation for research in economic and social sciences at the international level. This urgently requires a great number of measures.

In particular, it is necessary to develop and pass on to the scientific community the databases that exist at Eurostat and are harmonized for the EU Member States, including the national Scientific Use Files in the form of anonymized Scientific Use Files. The German Federal Government and statistical offices should strongly demand, support and enforce relevant measures.

The harmonization of data from different countries is in itself an urgent matter of research.

32. *The Commission recommends that the Federal Government – for the purpose of supporting data exchange with research institutions in non-EU Member States, in particular the U.S.A., Canada, Australia, Japan – take up negotiations aimed at implementing a “Safe Harbor” mechanism.*

An agreement on reasonable data protection at research institutions in “third countries” in the form of self-regulation and self-certification by the scientific community is considered by the Commission as urgent and suitable for sustained improvement of data exchange and cooperation between empirical economic and social sciences.

Demand for services and service agency for microdata

33. *The Commission recommends that research service institutions in Germany be maintained also in the future as part of the informational infrastructure.*

The efficiency of using microdata for research purposes may considerably be enhanced if various basic activities do not have to be performed ad hoc by the individual researchers or research groups but are rather done continuously and systematically at central service institutions for the entire scientific community.

It is important that the service functions are organized and performed in close contact with the scientific community. This ensures that the service institutions can perceive at an early point in time scientific developments and the ensuing demand for data and services and that they can shape their data and service offers accordingly. At the same time, such institutions must be independent from data producers, so that they can observe from some critical distance the trends emerging there, concentrate the interests of the scientific community and represent them towards the data producers.

The Commission supports the improvement of such services for, among other things, better data documentation, the development of anonymization concepts for new data files that serve the needs of both users and data protection, improved user service and cooperation with data owners, the setting up of metadatabases, participation in the further development of data protection regarding data in the field of science and, to a limited extent, own research activities.

Data linkage

34. *The Commission recommends that legal provisions be developed on the possibility of exact linking of data for purely statistical purposes, without the explicit consent of every respondent being necessary. Generally, exact linking of survey data with register data should be treated in the same way and be governed by the same legal provisions, too.*

To reduce both the costs of data acquisition and the burden on respondents, the providers of statistical infrastructure should more than in the past focus their efforts on full utilization of the information content of existing data material through linking of microdata for statistical purposes.

A precondition of exact linking of data that is unobjectionable in terms of data protection is ensuring the anonymity of statistical units with regard to government action (administration and justice) and commercial interests. Therefore, matching of data files should be possible only in a completely shielded research and statistics area to be defined explicitly.

Confidentiality of research data

35. *The Commission recommends that the legislator introduces the legal principle of “research data confidentiality”.*

A legal principle of research data confidentiality (*Forschungsdatengeheimnis*) avoids the problem of difficult trade-offs in statistical law and in general data protection law regarding the “necessity” of processing specific data by government agencies for research purposes and makes delicate interference in the freedom of science unnecessary. Research data confidentiality must be shaped in a way that both the scientist’s privilege to refuse testimony as a witness with regard to research data and a prohibition of seizure (*Zeugnisverweigerungsrecht* and *Beschlagnahmungsverbot*) are laid down in the law. Also, research data confidentiality and a breach thereof should be included in the Penal Code (in Germany: *Strafgesetzbuch [StGB]*).

The Commission holds the opinion that research data confidentiality should be implemented in the second stage of adjusting national data protection law and other provisions of area-specific data protection to the European Data Protection Directive.