

The Dissemination of Official Microdata in International Comparison

Report for the Commission to improve the
informational infrastructure by cooperation of the
scientific community and official statistics (KVI)

Franz Kraus
Bernhard Schimpl-Neimanns*

February 2001

Translated by
Dr. John Bendix, CGS Program, University of Pennsylvania,
Philadelphia, Pennsylvania, USA (E-mail: johnbx@sas.upenn.edu)

**) in collaboration with Nadia Granato and Simone Schmidt (ZUMA); authors are listed
in alphabetic order*

Franz Kraus
Universität Mannheim / University of Mannheim
Mannheimer Zentrum für Europäische Sozialforschung
(MZES) / Mannheim Centre for European Social
Research
EURODATA research archive
D-68131 Mannheim
Voice: +49-621 / 181 – 2832
Fax: +49-621 / 181 – 2834
Email: Franz.Kraus@mzes.uni-mannheim.de
WWW: <http://www.mzes.uni-mannheim.de/>

Bernhard Schimpl-Neimanns
Zentrum für Umfragen, Methoden und Analysen
(ZUMA) / Center for Survey Research and
Methodology
Abteilung Mikrodaten / Microdata Department
P.O.B. 12 21 55
D-68072 Mannheim
Voice: +49-621 / 12 46 – 263
Fax: +49-621 / 12 46 – 100
Email: Schimpl-Neimanns@zuma-mannheim.de
WWW: <http://www.gesis.org/en/zuma/index.htm>

Table of Contents

Foreword

1	Introduction.....	1
2	The Scope of Data Accessibility.....	5
2.1	Public Use Files and Scientific Use Files	5
2.2	Brief Description of Data Access in Select Countries	6
2.2.1	France.....	7
2.2.2	United Kingdom	8
2.2.3	Canada	10
2.2.4	The Netherlands.....	12
2.2.5	Norway.....	13
2.2.6	Austria.....	15
2.2.7	U.S.A.	17
2.2.8	Denmark	19
2.2.9	Sweden.....	21
3	Cooperation between Official Statistics and the Research Community in Various Tasks in the Context of Data Dissemination	25
3.1	The Opening of Data Access and Anonymization.....	25
3.2	Financing	28
3.3	Data Dissemination.....	34
3.4	Expansion of Analytic Potential	37
3.5	Consultation Services Provided.....	39
3.6	Assessment of Data Access by Researchers	41
4	Researcher Access to Social Security Microdata.....	44
5	National Models and Solutions: a Comparative Synopsis.....	52
	References	58
	Appendix: Documentation of the survey of statistical offices and service institutions of the research community.....	68

Foreword

This text is based on two reports that were conducted for the ‘German Commission to improve the informational infrastructure by cooperation of the scientific community and official statistics’ [Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik or KVI]. This Commission was called into life in 1999 by Federal Minister for Education and Research Edelgard Bulmahn, and charged with the task of formulating recommendations to improve the cooperation between official statistics and the research community in Germany. Other countries were to be investigated to understand how they had addressed comparable data access questions. In this context the Commission asked Bernhard Schimpl-Neimanns for an update and extension of a previous paper (Schimpl-Neimanns/Kraus 1996).

The first report focused on institutions in selected countries, both the official statistics agencies and the service institutions available to or created in the research communities, and asked of each a series of questions about data access (Schimpl-Neimanns/Kraus 2000). One selection criterion was variation across the selected countries in the tasks and services provided by the respective institutions of each research community.

Augmenting this comparison was a questionnaire administered by KVI to users and providers of microdata about the feasibility of data access. The results of this survey form the basis of the second report submitted to the Commission (Granato/Schimpl-Neimanns/Schmidt 2000). This English translation of the report is based largely on the first report. Where the second report provided additional information it was integrated here, including the country reports on Denmark and Sweden (section 2.2.8 and 2.2.9) as well as the evaluation of data access from the user’s perspective (section 3.6).

We would like to thank the Commission for their support for this translation.

1 Introduction¹

In empirical research, the official statistical information that is provided in the form of microdata about individuals, households, or enterprises are of considerable importance as a data basis. Unlike the aggregate data that is often published in tabular form by statistical offices, microdata provide anonymous particulars about basic socioeconomic units and permit the features under investigation to be flexibly combined with complex modes of statistical analysis. Access by researchers in Germany to official microdata has improved considerably in the last few years due to a pilot project that has addressed the problems of cost. In addition, this pilot project has facilitated the discussion between statistical offices and the research community about how to improve cooperation and develop new means to process and prepare the data. Our report provides a comparative overview of the dissemination of official microdata in selected countries. The cases described are meant as suggestive models for future institutional forms data dissemination could take in Germany, in particular with respect to better accommodating research needs.

The advantages of microdata from official sources come to the fore when one compares them to data collected by the research community: Official data more often than not contains very large numbers of cases; response rates are very high as official surveys frequently involve an obligation to provide information; and, last not least, official data allows longitudinal investigations, due to relatively constant survey questions that stretch far into the past. Economic and social structures and processes can thereby be analyzed with a high degree of differentiation, making it possible for reliable statements to be made about even very small populations; it permits social and economic changes to be observed on a continuous basis. As a consequence of the design of household samples, the behavior of individual actors in the family context can also be investigated.

For the research community to be able to utilize this potential, access to such data in as simple a manner as possible is necessary. From an individual researcher's point of view, a precondition for full use is good documentation that shows which data are available for addressing various research questions. It would also be desirable to have minimal expense in

¹ This report would not have been possible to complete in such a short time without the answers provided to our questions by the various representatives of the statistical offices and data archives. We would like to particularly thank the following: Erik Austin, Jeanine Bustros, Keith Cole, Ron Dekker, Peter Findl, Irène Fournier Mearelli, Charles Humphrey, Michel Jacod, Vigdis Kvalheim, Günther Nemeth, Joris Nobel and Roxane Silberman. For additional and constructive suggestions on this report, we would like to also thank

time and money to acquire the data, to have free choice of appropriate analytic methods, and to find consultative support and advice available on matters of evaluation.

Only since the revolution in computer technology has it been possible for the research community to engage in its own, independent evaluation of such comprehensive individual data from official statistical sources. In the 1970s in Germany, empirical social and economic research, for example through the SPES-Project², was able to catch up to the developments in the USA, where statistical offices began making Public Use Files available to researchers already in the early 1960s. At the time, the relatively open rules on data access led to an intense utilization of official household samples (in particular, of the Microcensus as well as of the Sample Survey of Income and Expenditure), and led to numerous publications. A few years later, however, as a consequence of the intense public discussion of the need to protect the privacy of individual data, as well as the promulgation of more restrictive legal regulations, scholarly access to official microdata for those conducting secondary analysis became very restricted by the 1980s. Microdata from the statistical offices could only be provided at such high levels of aggregation or in such crude forms that they were inadequate to meet the evaluation goals researchers had set.

Only with passage of the Federal Statistics Act in 1987 did the basic conditions of access to official microdata improve. This law created the possibility of passing on factually anonymized microdata to the academic community by making access a “privilege granted to research”. Specific rules for making data of the Microcensus and the Sample Survey of Income and Expenditure confidential were developed as part of a joint research project by the Federal Statistical Office, the University of Mannheim, and ZUMA [Center for Survey Research and Methodology] (Müller et al. 1991), as were other technical and organizational protective measures. Though this clarified the practical aspects of transferring data to researchers, microdata from official statistical sources remained little used until 1996 owing to the very high costs individual researchers incurred for data processing and preparation.

In the last few years, the access by the research community to official microdata in Germany has been considerably improved with the help of a pilot project developed and conducted by the Ministry of Education and Research (BMBF), the Federal Statistical Office, and GESIS

Nadia Granato, Paul Lüttinger, Simone Schmidt and Heike Wirth.

² Project Sozialpolitisches Indikatoren- und Entscheidungssystem (Socio-Political Indicators and Decisionmaking System) of the Universities of Frankfurt and Mannheim, financed by the German National Science Foundation (DFG; 1972-78).

[German Social Science Infrastructure Service]. In this pilot project, the Ministry provides the financing for the basic cost of data preparation, thereby in effect removing a major barrier to the utilization of the microdata selected for the project.³ The sharp increase in the use of official microdata in Germany is a direct result of this pilot project, though the project is slated to end in 2002/2003 (Köhler et al. 2000).

Given the precarious state of funding for data collection, a thrifty use of the available resources is important. Secondary analyses can often lead to considerable savings of monies that would otherwise be needed for collecting new data. The use by researchers of official microdata can also lead to an improvement of the data and contribute to the furthering of analytic knowledge and results. The varied possibilities for analyzing what have been complete surveys or very large samples conducted by official statistical agencies at considerable public expense have not nearly been exhausted by the standard analyses the statistical agencies themselves have carried out. Access to these data is part of the basic precondition of research in a democratic society. As with the data researchers themselves collect, official microdata should be accessible to the research community as well so that replication studies or reanalysis can be carried out. As a contribution to the advancement of knowledge, competing interpretations or critiques can thereby be offered. For official statistical data, this last point is even more significant, as official data are collected and used to a particularly high degree in the planning and justifying of political and administrative decisions (see Alba et al. 1994). The Panel on Confidentiality and Data Access in the USA came to the conclusion “that government dissemination of statistical data under appropriate confidentiality constraints is a public good“ (Duncan et al. 1993: 31), and based their arguments on the high value placed on official information, including that from official statistical sources, the importance of access to such information, and the personal right informants had to such access. They also suggested various procedures the USA could adopt in distributing data to the public.

Data privacy and financing problems can be observed in other countries as well. Since the late 1960s, data privacy and data access issues have been publicly discussed in many

³ This applies to the following data: Deutsche Stichprobe des Europäischen Haushaltspanels (German Samples of the European Community Household Panel) 1994-96, Einkommens- und Verbrauchsstichprobe (Sample Survey of Income and Expenditure) 1993, the 1995 and 1996 microcensus, as well as the 1991/92 German Time Budget Survey. To expand this data selection under the same terms and conditions of use, GESIS has agreed to finance the basic costs for the preparation of the 1989, 1991, 1993 and 1997 microcensus data.

industrialized countries. In certain countries, solutions have been found for the tension between data privacy and the right of self-determination over information on the one hand, and the research need for access to official data sources on the other hand. The problems of high data costs are also not specific to Germany. Bottlenecks in national budgets have led in some countries to changes in the fees statistical offices charge when they supply microdata to researchers. Comparative information is useful not only for improving the cooperation between empirical research and official statistical agencies in the microdata transfer itself, and in finding practical processes that are in line with research needs, but also because it indicates how problems have been resolved in other contexts and which procedures for disseminating and utilizing data have been developed.

Because the transfer of official microdata depends on the respective legal framework and institutional conditions in each country, conditions whose diversity makes them difficult to describe in detail, it seems sensible to concentrate on only a few select countries: France, United Kingdom, Canada, the Netherlands, Norway, Austria, USA, Denmark and Sweden. This limitation makes it easier to describe what is specific to each case.

Most of these countries can look back on lengthy traditions of statistical offices supplying data, as well as on well-developed data infrastructures for research, especially for users. Special agreements have been reached between the statistical offices and the research community in particular in Great Britain, the Netherlands and Norway, permitting institutions that are part of the research community to assume functions in the dissemination of data that otherwise ordinarily stay in the hands of the statistical offices.

With the exception of Canada, the Netherlands, Denmark and Sweden, access to official microdata is possible not only through the statistical offices but also through the institutions organized within the research community itself. In Canada and in the Netherlands, special microdata agencies were established in conjunction with the resolution of cost and financing questions. With respect to the question of what institutionalized forms cooperation between official statistics agencies and the research community takes, the tasks of data archives entrusted with the passing on of data are therefore of particular importance.

In the following, second section, an overview is provided of the legal and institutional frameworks and conditions in the individual countries selected for study. In the third section, the functions that the statistical offices and the institutions providing services to the research

community are described.⁴ The results of the survey on the dissemination of data from the social security agencies are reported in the fourth section. A concluding contrast summarizes the most significant aspects of the country models.

In general, the information is derived from questionnaires submitted to the statistical agencies (or administrative bodies in the case of social security records) as well as to institutions of the research community. Additional information was provided through a survey of data users in the selected countries.

2 The Scope of Data Accessibility

Anonymized microdata, in the form of Public Use Files, have been made available by statistical offices since the 1960s. The pioneers were the USA and Canada, and such data has been used by numerous researchers to answer the most varied types of questions. The other form in which data are supplied that we address in this report are Scientific Use Files, though these are accessible to the research community only under specific terms and conditions of use. The focus here is on these two types of official microdata⁵ on persons and households. Such data are an important component of the data infrastructure used in empirical social and economic research.

2.1 Public Use Files and Scientific Use Files

A basic distinction can be drawn between the restricted and the unrestricted use of data. Restricted use includes a contractual specification of research purpose, the length of time data will be used, and which persons will be entitled to have access to the data.

Public Use Files are usually made available to anyone by the statistical offices, and have no restrictions as to their right of use. To ensure sufficient confidentiality for such unlimited dissemination, individual data must be made so anonymous that it is impossible to identify

⁴ More extensive information on the individual countries reviewed here may be found in the appendixes, which contains the answers provided by the statistical offices and the data agencies that provide service to the research community; both provide details about their organizational structures and the modes by which data is disseminated.

⁵ We can only make parenthetical reference to alternative or complementary modes by which official microdata could be used, for example through remote access or via on-line data analysis tools. Also not considered here are newly created possibilities for analysis in the statistical offices themselves or at special research data centers. Such possibilities might provide researchers access to the original data, or access to particularly sensitive microdata, such as that provided by business.

the basic units (for example, persons or households) again. This is frequently simply referred to as “absolutely anonymized data“. Because of the need to protect privacy, coarsening of the data has to be undertaken, thus limiting the possibilities for analysis at least compared to the original data.

To make data more usable for the research community, additional protective technical and organizational measures can be undertaken in addition to those disclosure limitation methods that are specific to the data file (for an overview, see Jabine 1993; Müller et al. 1991; Willenborg/de Waal 1996). The preparation of microdata under contractually agreed-upon terms and conditions of use (**Scientific Use Files**, microdata under contract), as practiced in other countries is generally similar to the possibility noted in Art. 16, par. 6 of the German Federal Statistics Act of 1987. That paragraph stipulates that factually anonymized individual data can be supplied to the research community if the particulars of those surveyed can only be (re)identified after a disproportionately large investment of time, cost, and labor. The “factual anonymity” thus provided is matched by the particular obligation laid upon recipients of the data to respect legal data privacy provisions.

Public Use Files and Scientific Use Files are the data sources researchers most frequently use, as they are standard products of statistical offices: they are generally inexpensive to acquire and can be evaluated with the technical analysis instruments (both hardware and software) available in the respective research environments.⁶

2.2 Brief Description of Data Access in Select Countries

How official microdata is supplied to academic researchers varies by country, and the specific characteristics in each case have developed out of differing legal frameworks and institutional preconditions. Since the third part of this report describes the various ways such characteristics affect the division of specific tasks, what follows next is a country-by-country overview of the legal basis for transferring data, the accessible data basis, and the current cooperation between official agencies and the research community. The most significant legal and institutional conditions are summarized in an overview table at the end of this section.

⁶ In addition, other access possibilities exist that, like Scientific Use Files, belong to restricted access to data: remote access, online access, masked microdata, evaluation in statistical offices as part of a stay by a guest scientist or as part of a fellowship program, and work at research data centres.

2.2.1 France

The system of official statistics in France is characterized by a decentralized structure, a relatively large significance attached to data collection through surveys, the connecting of survey data with register data if necessary, as well as the linking of data production with content research.

No overarching national legal basis exists to regulate public access to microdata from the official statistics. However, the National Data Processing and Liberties Commission CNIL [Commission National de l'Informatique et des Libertés]⁷ and the Data Protection Committee of the National Council of Statistical Information CNIS [Conseil National d'Information de Statistique] are two bodies that are empowered to make basic decisions about access to confidential microdata. The CNIL is responsible for basic questions about access to data on households and persons, while the CNIS regulates access to business microdata through its Data Protection Committee.

It is French practice to give great weight to the right to information. For researchers, the data offering is broad, and includes Census data, a range of cross-sectional surveys, and cohort data. A whole range of data from household and personal surveys⁸ are available as Public Use Files, and include labor force surveys and samples from the Censuses. Upon request and after anonymization, however, practically every survey can be supplied. The measures to secure confidentiality in most cases are limited to removing all direct identifiers, including identification by small geographical codes. Only in the case of the most recent Census data has the CNIL set particularly restrictive rules for the geographical codes.⁹ For geographically more detailed Census data, access requires permission from the CNIL on a case-by-case basis.

The fee structure imposed by the National Institute for Statistics and Economic Studies INSEE [Institut National de la Statistique et des Études Économiques] is based on a

⁷ Created by Law No. 89 on information technology, data storage and citizens' rights (6 January 1978). See <http://www.cnil.fr/>.

⁸ As a matter of principle, access to survey data from the commercial sector requires official approval. The CNIS Data Protection Committee makes the determination as part of the individual application process. In this area, therefore, only Scientific Use Files are made available. In principle it is also possible, under special conditions, to conduct analyses of non-anonymized data in restricted areas of the statistical office, as part of middle to long-term cooperative efforts with INSEE.

⁹ Against this background the researchers questioned in the KV1 user survey expressed their satisfaction about the insignificant loss of information regarding anonymization and the access to the data (information provided by Roxane Solberman [LASMAS-IdL, Paris] and Professor Dr. François Bourguignon (Département et laboratoire d'économie théorique et appliquée [CNRS-DELTA], Ecole Normale Supérieure, Paris).

calculation of marginal costs, limited to the recapturing of the costs of anonymization, documentation and distribution. Data distribution is direct. The scope of the national social science data archive CIDSP/BDSP¹⁰ is currently still limited to microdata from academic survey research. However, in the LSMAS-IdL [Laboratoire d'Analyse Secondaire et de Méthodes Appliquées à la Sociologie - Institut du Longitudinal] laboratory of the National Center for Scientific Research CNRS [Centre National de la Recherche Scientifique], one finds a data center specializing in official microdata. This center provides extensive infrastructure, ranging from microdata purchases and documentation through advice on methodology and content, though restricted to CNRS institutes or associated university research institutes. The Ministry for Education, Research, and Technology [Ministre de l'Éducation Nationale, de la Recherche et de la Technologie] contracted a study of this situation (Silbermann 1999), as this arrangement excludes a significant portion of graduate educational institutions (écoles doctorales) as well as professional research, and on the basis of its conclusions suggested solutions be worked out to ensure a broader supply of official microdata to the research community. Among other suggestions, the possibility of remote access¹¹ to personal data, the creation of restricted areas at INSEE for purely academic projects, and the involvement of microdata agencies¹² that would be part of the research community and which would acquire, prepare, document, and supply official survey data to scholars, were discussed.

2.2.2 United Kingdom

In the United Kingdom, the institutional peculiarities lie in the decentralization of official statistics, the fragmentation of the legal basis for statistics, a research infrastructure in the social and economic sciences that is unique in Europe, as well as a traditionally close cooperation between statistical offices and academic scholarship. Relative to Germany, the U.K. is particularly interesting because a relatively loose regulatory framework, augmented by behavioral and procedural codes, is coupled with widespread use of official microdata by researchers.

¹⁰ Centre d'Information des Données Socio-Politique/Banque de Données Socio-Politique. CIDSP/BDSP is, as is LASMAS-IdL, an infrastructural institution of the CNRS.

¹¹ With respect to the KVI survey addressed to data suppliers, INSEE offers the possibility of data utilization via controlled telecomputing. Apart from a subscription fee of 1,000 Euro annually, no further utilization-dependent costs arise.

¹² Such a centre will probably be created cooperatively between LASMAS-IdL and CIDSP/BDSP

No overarching, national legal basis exists to regulate public access to microdata from the official statistics.¹³ However, one can find an implicit privilege accorded to researchers enshrined in the Code of Practice,¹⁴ a document which transforms a multiplicity of individual legal determinations into procedural rules.¹⁵ These rules and regulations specifically permit the dissemination of anonymized information to governmental organizations and bona fide researchers as long as no contradictory legal regulations or obligations exist (Government Statistical Service 1984: 5). As a consequence of the tightening of data privacy provisions after 1998, access to genuine anonymized official microdata in Great Britain is almost exclusively¹⁶ confined to bona fide researchers.¹⁷

The microdata available to researchers about persons and households is extraordinarily broad, and ranges from traditional cross-sectional surveys (Census, Labour Force Survey, General Household Survey, etc.) to the longitudinal surveys that are connected to Census data, such as the ONS [Office for National Statistics] Longitudinal Study¹⁸ and the JUVOS [The Joint Unemployment & Vacancies Operating System] Cohort Database.¹⁹

Scientific Use Files are supplied exclusively through the UK Data Archive²⁰, or its online analog, the Manchester Information & Associated Services [MIMAS].²¹ Other than the Census samples, data are provided at no cost from the Data Archive. The purchase of Census data is funded by the British research community and other granting agencies. Fees for end users are levied at most for documentation and for data carriers.

(information supplied by R. Silberman in the context of the KVI user survey).

¹³ The Data Protection Act, last amended in 1998, sets forth important principles to provide for the protection and confidentiality of personal information. Various court rulings related to individual surveys, following the English case law tradition, have also set legal precedents.

¹⁴ See the Government Statistical Service 1984. The Code is being revised, as the 1998 Data Protection Act altered a number of key requirements.

¹⁵ “If proper safeguards on confidentiality are applied, it is to the general advantage if data collected for statistical purposes inside Government are also available to outside analysts and researchers” (Government Statistical Service 1984: 4).

¹⁶ The Labour Force Survey is currently offered online as a Public Use Microdata File through SPSS MR (formerly Quantime).

¹⁷ The access to sensitive data, as contained in the ONS Longitudinal Study, is permitted to qualified researchers in restricted areas of the Statistical Office. Remote use of the data, under secure conditions, is supported by the British research community through a service institution it financially supports. See <http://www.cls.ioe.ac.uk/>.

¹⁸ See <http://www.cls.ioe.ac.uk/Ls/homepage.html> .

¹⁹ See http://www.statistics.gov.uk/themes/labour_market/surveys/juvos.asp . Compare this with the Cohort Study financed by the research community. See <http://www.cls.ioe.ac.uk/Cohort/cstudie.html> .

²⁰ See <http://www.data-archive.ac.uk/> .

²¹ See <http://www.mimas.ac.uk/> .

The British academic community makes extensive microdata research support available, in a breadth that is unique in Europe. In addition to the UK Data Archive and MIMAS, there is also the Cathie March Centre,²² the Centre for Joint Longitudinal Studies,²³ and the Centre for Applied Social Surveys (CASS).²⁴ In these institutions, exemplary data documentation is produced, user groups are aided, and feedback about user experience is collected and communicated to the statistical office. Among other things, these institutions are also involved in developing additional methods for making data confidential (Dale/Elliot 1998). The access to Scientific Use Files is very inexpensive for the researcher. The British data dissemination practice is judged as good to very good by the researchers surveyed.²⁵ All of this – making data accessible to the research community on the part of the statistical office and the fostering of a data infrastructure on the part of the British research community – has led to a situation in which 7 out of the 10 files most frequently provided through the Data Archive come from the statistical offices (ESRC 1998).²⁶ The cooperation between official statistics and the research community is very close. It stretches from regular informational meetings about official statistics, organized by the Royal Statistical Society and the UK Data Archive, to consultations with social scientists about new surveys, and from the modification of existing instruments (including the Census) to the joint conduct of surveys as part of contract research (see Marsh 2000, among others).

2.2.3 Canada

Statistics Canada, which as the central statistical agency is responsible for most of the surveys conducted by state agencies, already made microdata available as Public Use Files in the 1970s. The statistical office conducts its own, varied research on methods, and in cooperation with the research community has begun to develop a series of innovative statistics for social statistics purposes (Fellegi/Wolfson 1999). However, owing to stringent data protection provisions in statistics law, this work is not available to researchers. Until the

²² See <http://les.man.ac.uk/ccsr/Default.htm> .

²³ See <http://www.cls.ioe.ac.uk/Research/jclr.htm> .

²⁴ See <http://www.scpr.ac.uk/cass/> .

²⁵ Information provided by C. Mills and H. Sutherland in the context of the KVI user survey. H. Sutherland suggests a more effective distinction between scientific research and private research support. The aim should be to have different utilization levels according to the required or desired level of confidentiality. Furthermore, it should be possible to copy and use Public Use Files without permission for teaching purposes, for example.

²⁶ The data archive provides references to all relevant scientific publications for every survey. See <http://biron.essex.ac.uk/cgi-bin/biron/> .

creation of the Data Liberation Initiative (DLI), high financial barriers existed in the 1980s for researchers wishing access to Statistics Canada's Public Use Files. Against the backdrop of a similar problem constellation in Germany, the Canadian model is thus of particular interest.

The Statistics Act²⁷ was revised in 1971, and since then most social statistics surveys in Canada have been available as Public Use Files, including Censuses from 1971 to 1996, the Survey of Labour and Income Dynamics, the Work History Survey from 1981 and historical Public Use Files.²⁸ Aside from the Public Use Files, a variety of other data can be used, though as they involve special preparation they also carry additional costs. Researchers can also gain access to survey or administrative and register data which are considerably more data-rich than the Public Use Files, but for this they must conclude special contracts as to the terms and conditions of their use.²⁹

Until the early 1980s, Public Use Files were available and inexpensive, as fees charged covered only the cost of making the data ready for use. But a change in the fee structure by Statistics Canada made costs rise so sharply that not a single research institute could afford to acquire the microdata from the 1986 Census.³⁰ To resolve the problem of high costs, a financing model was subsequently developed that foresees a partial recapturing of the data costs through subscription fees. An independent data dissemination office within Statistics Canada was also established. The local DLI contacts at the universities provide the liaison to researchers. The External Advisory Board of the DLI has an equal number of representatives from the research community and from Statistics Canada.

The DLI project has led not only to a sharp increase in the use of data by researchers, but it has also been key in improving the cooperation between official statistics and the research community. From the viewpoint both of researchers and Statistics Canada, however, a number of problems remain. On the research side, the lack of a national data archive means the

²⁷ See <http://www.statcan.ca/english/about/statact/html> . The Corporations Returns Act also states which statistics are to be made available. In addition to the Statistics Law, which stipulates strict protection of data for information about individuals, the Privacy Act and the Access of Information Act also contain data protection regulations and conditions for access to the data.

²⁸ See <http://www.statcan.ca/english/Dli/list.htm> .

²⁹ With respect to the analysis of Public Use Microdata Files there are limitations in a general sense due to the coarsening of certain information, which makes it impossible, e.g., to analyze age-specific samples of the consumption and savings behaviour and to carry out certain politico-economic evaluation studies dealing with the implications of dissimilar provincial legislation (information provided by Professor Michael Veall in the context of the KVI user survey).

³⁰ Based on a governmental proposal, the fees for supplying data for Public Use Microdata Files have since partly supported the costs of data collection. In this, no distinction was made between research use and

data infrastructure is both inadequate and incomplete. On the supply side, there is also the serious limitation that at the moment the only access to microdata that is available to researchers is to Public Use Files. To make microdata containing more information available to researchers, a new initiative has been launched by Statistics Canada and the Social Sciences and Humanities Research Council (SSHRC) that proposes the establishment of research data centers at various universities.³¹ To modernize the data infrastructure for researchers, generous funding has been made available (Final Report 1998; SSHRC 1998).

2.2.4 The Netherlands

In the Netherlands, official statistics are completely centralized and Statistics Netherlands [Centraal Bureau voor de Statistiek, CBS] is in essence responsible for all statistical surveys and results. The statistics program established by the CBS must be approved by the independent Central Commission for Statistics [Centrale Commissie voor de Statistiek; CCS]. Since the 1970s, Statistics Netherlands has increasingly conducted its own research and has been particularly concerned with developing methods for statistical disclosure control. As in Canada, the solution to data access problems engendered by high fees has been provided by the development of a special financing model. A data agency anchored in the research community has been founded, its costs to be recouped through fees for acquiring the data.

For social statistics, the CBS has a variety of datasets that are of great interest to researchers, including the Socio-Economic Panel, the Mobility Survey and the Permanent Study on Living Conditions.³² Because of public resistance to conducting the Census already in 1971, no complete Census surveys have been conducted since that time (de Vries/Nobel 1999). Statistics Netherlands has relied more on register data since, and has utilized various other sources of data.

private use (such as by banks). See Albert et al. 1996; Statistics Canada 1997.

³¹ The analysis centre that was newly established at the McMaster University, Hamilton, provides a secure area where researchers can work with selected original data from Statistics Canada (Master Files). See also Statistics Canada Research Data Centres (2000): Cf. Handbook for Researchers Under Agreement with Statistics Canada (Draft – October 2000). WWW-Site: <http://www.socsci.mcmaster.ca/rdc/rdcinvit.htm> ; 26.10.00).

³² Compare this with the WSA data agency at http://wsa.gamma.rug.nl/index_uk.htm .

For social statistics surveys, researcher access to data is regulated by the Statistics Law of 1996,³³ as in Germany for data rendered anonymous. Dutch law is more open, however, inasmuch as research institutions outside of universities, including state planning agencies or other certified state research institutions, can obtain access to the data if they meet certain criteria.³⁴ Microdata under contract can be provided for research, under particular contractual stipulations as to use, but only in the form of samples from the social statistics.³⁵

Once a contract has been concluded between a research institution and the CBS, the Scientific Statistical Agency WSA³⁶ [Wetenschappelijk Statistisch Agentschap] functions as the coordinating body for the dissemination of all microdata provided directly by Statistics Netherlands. The WSA itself owns no data (Nobel 1995). User support in the Netherlands is carried out both by Statistics Netherlands and by the WSA. In the view not only of Statistics Netherlands but also of the research community, the resolution of cost issues and the creation of the data agency has led not only to a considerable increase in data utilization but has also strengthened cooperation between the Statistics Netherlands and researchers.³⁷ To make more CBS data sources available, new data access procedures will be tested in cooperation with the research community.

2.2.5 Norway

What is typical for official statistics in Norway, as in other Scandinavian countries, is the extensive use of administrative sources and its connection to survey data, the strong research orientation of the statistical office, the extraordinarily stringent levels of data protection (Bing 1996; Kvalheim 1999), the privileged access of researchers to data, and a level of cooperation between official statistics and scholars in the preparation of official microdata that is probably the closest in Europe. At first glance, Norway seems to demonstrate an odd coexistence of high data intensity, with many connected data from administrative sources and stringent data

³³ Also relevant are the general Data Protection Act (1988) and the law on the Census of Business (1936) (see de Vries/Nobel 1999).

³⁴ See the comments of Joris Nobel (CBS) in Appendix 4A.

³⁵ In order to make business data, which cannot be provided as microdata for legal reasons, accessible to researchers, in 1998 the creation of two analysis centres within the CBS was started. This project is financed through funds of the Dutch research agency NWO (information supplied by the CBS in the context of the KVI survey of data suppliers).

³⁶ The statistical office has a seat on the advisory board of the WSA data agency, though it has no voting powers.

³⁷ In 1999 there were 71 orders for data coming from 35 research organizations. An estimated 1,000 individual researchers used the data in 1997 (see information supplied by Ron Dekker (WSA) in appendix

protection, with very open data access accorded to researchers, but these very features indicate that the Norwegian case is particularly interesting.

Unique personal identity numbers permit the linking of different data sources, so Statistics Norway has an extraordinary instrument of information-gathering at their disposal. The range of information includes cross-sectional surveys, cohort studies and panels, and includes complex data banks containing extensive event history data. Norwegian official statistics thus offer diverse possibilities for scientific analysis not only in terms of content but also in methodological terms.

The 1989 Statistics Law restricts access to official microdata to researchers and public planners.³⁸ The form access takes is not determined by the statistical office itself but is instead the responsibility of the data inspectorate,³⁹ an institution that not only acts as a superordinate body to oversee data protection but also regulates the statistical office itself. In the license it granted the statistical office,⁴⁰ the data inspectorate set forth two central rules for access by the research community: obligatory official authorization by the inspectorate, on a case-by-case basis, for all data that was not absolutely anonymized, and an obligation on the part of the statistical office to limit anonymization to the necessary minimum in the interest of research.⁴¹

As a result, in Norway microdata can only be supplied in the form of Scientific Use Files.⁴² On the other hand, the data inspectorate has gradually moved toward a pragmatic solution to simplify the authorization process for bona fide researchers. The inspectorate has agreed to let the Privacy Issue Unit within Norwegian Social Science Data Services [NSD] handle approvals, and though the NSD is itself part of the research community, it acts as trustee for the statistical office (Kvalheim 1999: 145). As part of this agreement, the statistical office also designated the NSD as the institution that supplies official data to researchers.

4B).

³⁸ See http://www.ssb.no/english/about_ssb/statlaw/statlov_en.html .

³⁹ The tasks of the data inspectorate are set forth in Act No. 48 of 9 June 1978 Relating to Personal Data Filing Systems (see <http://www.datatilsynet.no/arkiv/engelsk/LOV-ENG.html>).

⁴⁰ In Norway, every person or organization wishing to obtain and store individual-level data must obtain a license to do so from the data inspectorate, regardless of whether this is for private commercial, scientific or governmental use.

⁴¹ The data inspectorate stipulates that individual data must be made anonymous, but only to the extent that such anonymization is congruent with the goal of the research (Kvalheim 1999: 142).

⁴² Those entitled are not limited to bona fide researchers but include all research institutions, including commercial ones.

The actual access by researchers to microdata on households and persons is comprehensive and efficiently organized.⁴³ Applications are processed by the Privacy Issue Unit in the NSD, and the NSD, as trustee for the statistical office, signs the contracts with the applicants wishing access to the data. Among other things, these agreements contain a written obligation on the part of the data recipients to abide by the data protection guidelines, and also limits the length of time the data may be used.

NSD receives the microdata from the statistical office at low cost. NSD services are publicly financed, so end users need only pay minimal fees to cover the costs of documentation and for data carriers. User support is provided both by the NSD and by the specialized offices in the Statistics Norway. The NSD also provides other services such as implementing additional statistical disclosure limitation methods following statistical office stipulations, cataloging of all surveys, creating data documentation in accordance with international standards, providing various consultation services, and securing the transfer of information between end users and the statistical office.

The close cooperation, which is based on consensus and the creation of trust, has not only relieved the statistical office from providing extensive services to researchers, but has also led to common projects, as in the creation of the Kirut data bank⁴⁴ with its complex event history data on social security, and the creation of a data bank on health research.

2.2.6 Austria

Since the beginning of 2000, the Austrian Central Statistical Office has been made independent of the federal branch and transformed into an institution under public law called Statistics Austria. Within the system, Statistics Austria remains centrally responsible for federal statistics. In the new Federal Statistics Act,⁴⁵ statistical findings are characterized as a public good, which among other things is underscored by an obligation to publicize and freely disseminate statistical publications over the Internet. A newly created Statistics Council is meant as a supervisory board to guarantee the quality of statistics, as well as ensure their

⁴³ Business microdata can only be obtained through the statistical office.

⁴⁴ The Kirut data bank is based on the linking of survey, register and other administrative data and was created by the NSD, working in close cooperation with the statistical office and the social welfare insurers. It was created with the permission of the data inspectorate (Kvalheim 1999).

⁴⁵ See the text of the act at <http://www.oestat.gv.at/institution/bundesgesetz.pdf>. For the Federal Statistics Act the following articles on the dissemination of data are particularly relevant: Bauer/Hackl 1999; Engelage 1999; Kutzenberger/Richter 1999; Münz 1999.

independence, and reflects efforts to increase the scientific character of statistical work.

In Austria, many microdata are accessible to researchers. They include the Census of Buildings and Housing, the Population Census, the Microcensus, and the Consumption Survey (see appendix 6A). Because the microcensus supplements (including retrospective surveys of employment history) were carried out without an obligation on the part of respondents to supply information, no special regulatory framework was necessary.⁴⁶ In this respect, the Austrian legal framework provides for more structural latitude than does the German.

In conjunction with the new Data Protection Act, the Federal Statistics Act [BStatG] has strengthened data secrecy in its effort to protect personal information. For personal data, researchers are not accorded any special position or privileges. In addition to the general obligation to give information, which foresees that in publications no possible reidentification of those surveyed will be possible, Statistics Austria can generally make anonymized microdata available for research purposes. Following Par. 31 BStatG 2000, these data must be used in a manner tied to the project or stated purpose. The measures to ensure data anonymity include erasure of names and addresses (inasmuch as they were collected), as well as regional identification of the individual community, though district-level and larger geographical units remain identifiable. Microdata in the form of raw data and written data documentation are accessible through Statistics Austria after a contract setting forth terms and conditions of use has been signed. The Federal Statistics Act sets forth that only additional costs may be calculated in.

Since the 1970s, anonymous microcensus data have been supplied to the research community at a moderate cost. In its conceptualization of the microcensus supplements, Statistics Austria continues to work closely with the research community; researchers are included, for example, in ad hoc working groups. Statistics Austria has thus been able to provide very interesting data to the research community, though owing to resource limitations was able to do its own evaluations and analyses only in very limited ways. The new directors intend to soon change this situation.

⁴⁶ However, a financing problem has been created by separating Statistics Austria from the federal administration. Special programs that are not part of the obligatory statistical tasks will need to find other public clients to pay for them (P. Findl/Statistics Austria; additional information provided on 16 May 2000).

The WISDOM data archive, founded in 1985, receives anonymized microdata from Statistics Austria and can pass these on to researchers in user-friendly formats following unbureaucratic procedures that meet the practical needs of scholarship. The cooperation between official statistics and the research community is judged to be good. Still, those at WISDOM are unsure about an anticipated increase in costs which may result from the decoupling of Statistics Austria from the federal administrative structure.

2.2.7 U.S.A.

Statistical information in the USA is strongly decentralized. Various official agencies and federal departments have their own offices which, based on independent legal mandates, conduct their own statistical investigations and report results and data.⁴⁷ From the research point of view, the most important federal statistical agencies include the Bureau of the Census, the Bureau of Labor Statistics, the National Center for Education Statistics, the National Center for Health Statistics, and the Social Security Administration. Noteworthy for the entire mode of statistical work is an understanding of democracy in which access to data is given the highest priority. Statistical offices in the USA began to make their microdata available as Public Use Files already by the early 1960s. Though access to these data is easy and the data itself is available at a very reasonable cost, the Inter-university Consortium for Political and Social Research (ICPSR) data archive plays a central mediating role for researchers.

The legal bases are anchored, among other things, in the Freedom of Information Act obligating all official agencies to make public all data that by law has not been explicitly excluded. The legal stipulations for the individual surveys, as well as in the Privacy Act, define the data protection rules. In Census Bureau statistical surveys, personal data can not be supplied, and strict confidentiality is usually observed. However, there are various ways of gaining access to such microdata, for example through fellowship programs or in research data centers,⁴⁸ that exist inside statistical offices as well as at the universities. If it can be

⁴⁷ The Statistical Policy Office in the Office of Management and Budget is responsible for the planning of statistical work and for setting the framework conditions as well as for coordinating various subunits and tasks.

⁴⁸ In addition to the analysis centre at the Bureau of the Census, which has existed for a longer time, further Research Data Centres that are linked to universities have been established lately (see <http://www.census.gov/cecon/www.rdcannounce.html>). In this context it is worth noting that research carried out at these analysis centres by external researchers has contributed to the improvement of official databases (de Vries 2000). Within the framework of the KVI user survey Professor Martin David

guaranteed that in the relevant anonymized datasets no reidentification of those surveyed is possible,⁴⁹ microdata can be prepared as Public Use Files, usually at a geographic aggregation level of regional units with 100,000 inhabitants. Particular information-rich data are only accessible under special terms and conditions of use and following registration.⁵⁰

A variety of microdata, usually as Public Use Files, are available from the various statistical offices, including Census data, the Current Population Survey, the Survey of Income and Program Participation, the National Education Longitudinal Survey, and New Beneficiary Data.

Public Use Files are accessible through the statistical offices at moderate cost, and the most important data have been made available for free on the Internet in the last few years, either as complete sets or as partial data. They are not only available through ICPSR, whose comprehensive infrastructure and services are detailed below, but can also be accessed in many university data libraries as well as through commercial information providers.⁵¹

As can be seen not only in the example of the free dissemination of 1990 Census data to the ICPSR (Austin/Rockwell 1997), the statistical offices in the USA support researchers' access to data in many ways. Numerous forms of cooperation exist with the research community, including that statistical offices are members of the Association for Public Data Users, and that they provide financing for fellowships and research data centers together with the American Statistical Association and the National Science Foundation. A variety of investigations are carried out together, as in the cooperation between statistical offices and the research community in surveys like the Panel Study of Income Dynamics. What is a matter of course on the part of the statistical offices in providing researcher access to microdata is matched by political support for the official statistical agencies on the part of the research community. Thus, in the contentious issue of the proper statistical resolution for a claimed

(University of Wisconsin, Madison) points out that the research project forming the basis of the application for data utilization must be of interest to the Bureau of the Census. He thinks that the amount of time needed for data analysis at the analysis centre is very high.

⁴⁹ For specific criteria for dissemination, see the Statistical Policy Office 1999; Subcommittee on Disclosure Limitation Methodology 1994; Zayatz et al. 1999.

⁵⁰ Users must swear an oath confirming that they will hold to the agreed-upon terms and conditions of use. These include: use solely for the stated research project, time limit on use, technical and organizational protection of the data received, and limitation as to the persons accorded access to the data.

⁵¹ Some researchers are willing to pay the higher prices commercial providers demand, as such data often contain additional information or have added value. Other useful tools may be provided that make data extraction and documentation easier. The Unicon Research Corporation (<http://www.unicon.com/>), for example, provides an array of tools for working with the Current Population Survey.

undercount in the 2000 Census, and in support of scientific modes of work independent of political demands, the American Sociological Association was willing to publicly defend the work of the Census Bureau (see Levine 1996).

2.2.8 Denmark

Danish statistics are generated in a highly centralized manner by Danmarks Statistik (DS), and since 1978 official statistics have been based primarily on data from public records. Administrative and statistical registries can be linked for statistical purposes, though an independent data protection authority is responsible for public record data as well as for linkage. Owing to budgetary constraints, data acquisition and preparation costs are largely covered through fees.⁵²

Making microdata available is one of the basic tasks of the statistical office. In addition, the national health service, individual communities, and the central bank also furnish microdata. In addition to the microdata from the statistical office, administrative and register microdata from Ministries are also in principle usable for research purposes. Examples of such microdata include public record data on the incidence of cancer, the central registry of persons, and the death register. The University of Southern Denmark also keeps a dataset that registers all twins born in Denmark.⁵³

The statistical office provides a variety of project-specific data files.⁵⁴ A Survey of Firms is available that is based on questionnaire data. Microdata files based on public records are also of great importance for labor market research. Among other available datasets, one can find longitudinal labor market data (since 1980 based on a sample of 5 percent of the population) and longitudinal employer-employee data. As one example of the many ways public record data have been linked, one can note the dataset that links data from the Longitudinal Labor Market administrative data together with data from the Psychiatric Case administrative data.

With only a few exceptions, no Public Use Files or Scientific Use Files are supplied to researchers in Denmark. It is generally only possible to evaluate Scientific Use Files in the

⁵² According to the statistical office answers to the KVI data producers questionnaire, income from the various scientific, governmental, and commercial users amount to around one-third of the total expenditures.

⁵³ Information provided by Prof. James Vaupel (University of Southern Denmark at Odense) in response to the KVI user questionnaire.

⁵⁴ Information provided by Prof. Niels Westergaard-Nielsen (Centre for Labour Market and Social Research (CLS), Aarhus School of Business) to the KVI user questionnaire.

statistical office itself, and to this end, Danmarks Statistik has data analysis centers in Copenhagen and Aarhus.

However, Public Use Files have been created recently which contain some of the variables employed in the labor force surveys. The results of the work in the statistical office, for example as presented in tables or in other aggregated forms, can be used by researchers outside the data analysis centers once their adherence to data protection rules has been ensured. Administrative data, in the form of special analyses conducted by the Ministries or statistical offices, can be utilized, though they are not free and the disseminated results must also meet certain data protection criteria, such as a minimum number of cases per cell in a table. In contrast to the decidedly guarded approach of Danmarks Statistik in providing microdata to the research community, these kinds of data are available directly, under certain conditions, from Ministries and the state pension system. Taking this route requires the approval of the data protection offices, but according to researchers it has often been possible in this manner to acquire Scientific Use Files for medical research.⁵⁵

Evaluations carried out in the data analysis centers of the statistical office incur fees for the preparation of data and for computer use. 94 researchers involved in 62 projects made use of this service in 1999, according to the statistical offices, and the total fees amounted to about 800,000 Euro. There are no special terms granted for use by the research community.⁵⁶

According to the KVI user survey, researchers assessed the quality of the data documentation provided by the statistical office as good, but this was less true for data documentation from the Ministries. With regard to the easy access to public record data and its ready linkage to other data, Danish researchers evaluate both data availability and data access very positively. However, researchers also noted the limitations on use that high data costs impose and the exclusive control over access that the statistical office exercises.

⁵⁵ However, Danmarks Statistik does provide microdata to researchers in individual cases. The Center for Labour Market and Social Research (CLS, Aarhus) was able to acquire a continually updated 0.5% sample of the Longitudinal Labour Market administrative data from the statistical office. These data may also be used by other researchers at CLS (information provided by N. Westergaard-Nielsen).

⁵⁶ According to Prof. Vaupel, costs for computer use in the statistical office are modest. Prof. Westergaard-Nielsen also noted in the user survey that the CLS Longitudinal Employer-Employee administrative data can only be used for a limited time in the data analysis centers of the statistical office. In addition to preparation fees, user-dependent costs are also charged.

2.2.9 Sweden

Sweden has a decentralized statistical system in which 25 other governmental offices in addition to Statistics Sweden (SCB) perform statistics tasks. The statistics office itself is responsible for coordination and supervision of the entire statistics system, as well as for the further development of methodology, but will conduct its own surveys for individual governmental offices and private companies as well. Such surveys must have their costs reimbursed by those who contract them, and about half of all statistics are carried out under the auspices of the government statistics office. Since there are no legal mandates giving SCB the exclusive right to conduct surveys, and Ministries can contract surveys out to other institutions, the SCB finds itself in constant competition with other providers of survey data. As in Norway, oversight over all individual data lies with a data protection agency which issues licenses for conducting surveys.⁵⁷

Access by researchers to official microdata is regulated by the Secrecy Act (1980, 1994) and in the Data Act (1973, 1994).⁵⁸ Most universities are part of the public sector, so researchers are subject to the same data protection rules that statistical offices are subjected to as well.⁵⁹ The statistical office can provide microdata for researchers for specific research projects.⁶⁰ One of the contractual terms stipulates length of use, and when that time has expired, the data must be erased or deleted. With the agreement of respondents, information from administrative data (for example income or tax data) can be passed on to the researchers.⁶¹

⁵⁷ From the SCB response to the KVI data producer survey.

⁵⁸ See the Handbook of Official Statistics at <http://www.unec.org/stats/handbook/legal/swe.htm> (accessed: 1 November 2000).

⁵⁹ Following the older Data Protection Law, only limited criminal prosecution was possible for violation of data privacy. Until the early 1990s, the statistical office was very cautious about providing data to researchers owing to a spectacular case of data misuse in 1986 called the “Metropolit Incident”, according to Prof. Klevmarken (Uppsala University). One consequence was also a subsequent reluctance by informants to provide information and answers to survey questions.

⁶⁰ Though there is a proper procedure for data inquiries, in Prof. Jonsson (Stockholm University) and Prof. Klevmarken’s opinion it is often easier to order the data directly from the relevant departments in the statistics office. In this context, J. Jonsson noted in his answers to the KVI user survey that outside researchers themselves are partly regarded as competitors by the specialized departments in the SCB, and that can lead to delays and obstruction in passing on the data. It is worth noting in this context that in Sweden, unlike for example in the USA, access to data produced by the research community is often equally difficult since researchers are not obligated at the end of projects to supply their surveys to the social science data archive even when the work is financed with public monies.

⁶¹ For example, according to Prof. Klevmarken, information from the administrative data, based on social security numbers, was provided for the Swedish Household Panel Survey Market and Nonmarket Activities (initiated by Klevmarken himself) as well as for the Level of Living Survey. In the Household Panel, about two-thirds of the respondents agreed to providing this information.

No Public Use Files are prepared in Sweden. According to the statistical office, however, Scientific Use Files about individuals and households, based on samples and administrative sources can be disseminated to university as well as non-academic researchers.⁶² Survey and register or administrative firm or business data can in principle be accessed, but the SCB did not provide more precise information on this possibility.

Much administrative data is available. Owing to an event-analytic design as well as the use of social security numbers, and owing to linkage possibilities with other administrative and register data, as well as with survey data, it is possible to construct complex data files containing information about income, taxes, skill levels (and educational careers), as well as about aspects of the social insurance system. Researchers can also use the complete surveys conducted by the National Insurance Board that contain information on income, pensions, social welfare support, and length of unemployment.⁶³ For epidemiological questions as well as health inquiries, administrative data from the Center of Epidemiology are accessible.

Official microdata is supplied after a contract has been signed that details the project-specific and time-limited use, with microdata coming exclusively in the form of Scientific Use Files. The SCB provides data carriers in various formats (ASCII, SAS, SPSS, etc.) at cost.

From the point of view of researchers, the situation with regard to data access has improved in the last few years and is generally positively assessed, but significant problems remain. These include the exclusive access through the statistical office, the long delays in preparing requested data, the high costs of acquiring data, the lack of standard files, and the inefficient procedures used for updating already extant Scientific Use Files.⁶⁴ Owing to political constraints that are related to the high costs of data, it is worth noting that the statistical office must have its data commercially processed.⁶⁵

⁶² Labour Force Surveys, Household Expenditure Surveys, Time-use Surveys, Welfare and Level of Living Surveys, and the Survey on School Leavers are among the sample surveys of the statistical office that collect data at the individual and household level.

⁶³ Information provided by Prof. Dr. Nils Anders Klevmarken on the KVI user survey.

⁶⁴ The current interpretation of data protection by the statistical office does not permit storage of an identifier in the office itself. The result is that it is not possible to track information in the Scientific Use File back to the original file (for example, through an anonymized key of the social insurance number). To update existing files, therefore, the original material must be consulted again and a new file created, a process that is time-consuming and costly. In the case of LINDA, an event dataset constructed from various administrative data sources, this problem was resolved with the creation of a special financing model.

⁶⁵ For example, for the time-limited use of data files from the Survey on Living Conditions (with around 150,000 cases and 150 variables), the cost was around 25,000 Euro. The creation of a file with information about educational development and social background (with around 1 million cases) based on administrative data cost 40,000 Euro (information provided by J. Jonsson in the KVI user survey). In Prof.

In **conclusion**, the selected countries can be distinguished in an ideal-typical manner between those cases in which comprehensive legal frameworks set the tone and those countries which operate with individual laws and detailed implementing provisions governing data dissemination. Legal frameworks provide greater latitude for unbureaucratic data dissemination and problem-solving. What also varies considerably is how data protection is regulated; in some countries this is an important avenue of access to official data for the research community. In general, it is easier for the research community to reach agreements about the dissemination of data with statistical offices if there is a central statistical office that is responsible for all official data, at least compared with strongly decentralized statistical systems. However, it is also important if the statistical bureaus themselves are receptive to the needs and wishes of the research community. Statistical offices often see themselves as part of the administration, but owing to the strong research orientation within official statistics, it has been possible for cooperative relations with university-based research to develop with consequent good effects on data dissemination practices. Based on the goal of presenting cases where utilization and dissemination practices have been developed, in the countries examined here, the majority has statistical offices with long research traditions. In terms of the research community, most countries examined here have well-developed data infrastructures, usually reflected in the existence of a national data archive which holds data ready for secondary analysis, though there are nation-specific forms such infrastructures take.

Overview 1: Legal and Institutional Frameworks

Country	Statistics and Privacy Laws	Institutional Frameworks
France	Decentralized statistical system, Coordination through INSEE. Law 711 of 7 June 1951 (Framework law for surveys with obligation to provide information); Decree 84/628 of 17 July 1984 (Creation of CNIS); Ministerial decrees; Data Protection Law (1978).	INSEE traditionally has a strong research orientation (CREST); institutionalized cooperation with university-based research (research cooperation with CNRS). LASMAS-IdL (1986) data archive is only available to CNRS researchers.

Klevmarken's view, the increased dependence of the statistical office on its own revenues has certainly led to a greater openness to client wishes and to improved data dissemination. In his estimation, the high data costs were related to the monopoly the office possessed as well as their inconsistent application of marginal cost calculations.

Country	Statistics and Privacy Laws	Institutional Frameworks
United Kingdom	Decentralized statistical system; Coordination through ONS. No legal framework for the areas of responsibility of statistics; case law is determinative. Data Protection Act (April 2000).	Research orientation of the office and the Ministries; close content cooperation between official statistics, Royal Statistical Society and the data infrastructure of the research community. Developed infrastructure (including Data Archive (1967) and MIMAS); 1974 agreement OPCS – ESRC Data Archive to disseminate official microdata surveys.
Canada	Centralized statistical system. Statistics Act (1971): Dissemination of anonymized data, access to personal data only as a “deemed STC employee”; Access of Information Act and Privacy Act (1982); Corporations Returns Act.	Pronounced research orientation of the statistical office. Crown Copyright governs use of public use microdata files. Data dissemination fees include participatory data acquisition costs. No national data archive for researchers.
The Netherlands	Centralized statistical system. Statistics Law (1966): Dissemination of anonymized microdata (on persons, households) only for research purposes; general Data Protection Law (1988); Law on Enterprise Censuses (1936).	Increased research orientation by the CBS since the 1970s. Data dissemination fees include participatory data acquisition costs. Steinmetz data archive (1964), WSA Scientific Statistical Agency (1994).
Norway	Centralized statistical system. Personal Data Registers Act (1976) as well as Statistics Act (1989).	Strong research orientation of the statistical office. Data inspectorate decides over access to anonymized microdata. Agreements on data dissemination through the NSD data archive (1976, 1997).
Austria	Statistics Austria is centrally responsible for federal statistics. Federal Statistics Act (2000): privileged access granted for researchers to anonymized individual units; Data Protection Act (2000).	Statistical office independent since Jan. 2000 as an institution under public law. Statistics as “a public good”. Cooperation between Statistics Austria and researchers, for example in the microcensus special programs. WISDOM research data archive (1985)
USA	Decentralized statistical system. Freedom of Information Act (1966); Privacy Act (1974); access rules specific to survey and institution.	Pronounced research orientation of the statistical agencies; close cooperation with researchers: Association of Public Data Users, ASA/NSF/BOC Fellowship Program. Public Use Microdata Files are “a public good”. ICPSR (1962) provides a developed data infrastructure for research.
Denmark	Centralized statistical system. Act No. 196 concerning on Statistics Denmark (1966); The Danish Public Authorities’ Registers Act, 20 Sept. 1991.	The data inspectorate must approve the use of departmental data (which is widely used in medical research). Such data can be processed on the premises of the researcher.
Sweden	Partly decentralized statistical system. Statistics Act (1992); Data Act (1994); Secrecy Act (1994).	Research orientation of the statistical office (particularly in the field of social indicators)

3 Cooperation between Official Statistics and the Research Community in Various Tasks in the Context of Data Dissemination

In the dissemination of microdata to researchers, a variety of tasks arise, from the construction of data files to data documentation to user support, whose provision is ensured in different ways in the various countries examined. In some cases, a complementary service is provided by the official statistics, while in others the tasks are undertaken through institutions established by the research community that act in a trusteeship capacity for the statistical office. In the following, based on the information documented in the appendixes, the different data transfer procedures and division of labor are detailed.

3.1 The Opening of Data Access and Anonymization

If the statistical offices prepare and provide Public Use Files or Scientific Use Files as standard products readily available to researchers, in most cases a division of labor and functions has already developed and been established between the respective statistical office and the institution in the research community providing user services. For the research community, having interests centrally addressed is both sensible and functional, as otherwise problems could arise owing to the fact that the data itself has the character of a public good. For the statistical offices, as the data producers, it is also an advantage to have one centralized institution to deal with on the researcher side. The Swiss Federal Statistical Office, for example, has written that “Pluralism in science means that research, which in the end organizes itself and has its own rules, multiple institutions, and organizational forms, is not a simple partner for official statistics to understand” (Haug/Buhmann 1998: 303).

In Norway and Great Britain agreements have been reached whereby the statistical office gives priority to data archives in its dissemination of microdata to researchers, or releases data exclusively through a data archive. In France, the LASMAS-IdL can only act for those research groups that work together with the CNRS, but in ongoing discussions between the INSEE and the Ministry for Education, Research, and Technology [Ministre de l'Éducation Nationale, de la Recherche et de la Technologie] the creation of new research data centers that will make official microdata ready and available to all researchers is under discussion (see the INSEE information in Appendix 1A).

Which investigations and surveys can be passed on to researchers in the form of anonymous microdata depends on the legal possibilities for doing so, but for various reasons these possibilities are not always fully realized. In this context, for example, it is part of the task of the Dutch Scientific Statistical Agency (WSA), in cooperation with the CBS, to present the data needs of researchers and to make new datasets accessible. In Canada, agreements about available data are reached in the DLI's External Advisory Committee, a body with equal numbers of representatives from Statistics Canada and from the research community. Regardless of the prevailing data acquisition model, however, some archives only become active after researchers themselves make known what their needs are (see Overview 2).

From the point of view of research, and beyond the palette of data offered, particular attention needs to be given to data utilization possibilities, since making data anonymous is necessarily connected with the limitations this places on analysis. In order that when microdata is made anonymous, those methods of disclosure protection are chosen which least limit later data utilization possibilities, it is important that researchers participate in drafting the planned microdata modifications for disclosure control. For example, in the Netherlands the project that led to the founding of the WSA began with the goal of working out an anonymization plan (for the dissemination of the Dutch Continuous Labour Force Survey) cooperatively between official statistics, data protection agencies, and the research community (Nobel 1995).⁶⁶

To provide researchers access to anonymized records of the 1981 British Census, an ESRC Working Group not only made suggestions as to file specifications, but in a larger project investigated the risks of possible (re)identification and suggested specific means to prevent it. Based on these efforts, since the early 1990s the statistical office has been able to provide researchers with samples of anonymized Census data, and make them available through the Census Dissemination Unit in MIMAS (Marsh et al. 1991; Marsh/Teague 1992; Marsh et al. 1994).⁶⁷ Based on the special agreement between the Norwegian research community, the

⁶⁶ The Dutch researchers questioned within the framework of the KVI user survey point out, however, that considerable limitations are placed on the data analysis potential due to the applied statistical methods of disclosure control (information provided by Professor Dr. Harry Ganzeboom (University of Utrecht) and Professor Dr. Ruud Muffels (University of Tilburg) in the context of the KVI user survey.

⁶⁷ In the context of the KVI user survey Colin Mills (London School of Economics and Political Science) points out that as a result of statistical methods for disclosure control geographical information cannot be included in the analysis in the form of macro information as it is too coarse. Professor Holly Sutherland (University of Cambridge) does not consider the loss of information due to disclosure control methods to be too severe.

statistical office and the data inspectorate, the NSD receives all the requested microdata from surveys on individuals and households, and is responsible for their dissemination to interested researchers as well as for further measures to render them anonymous. In the USA, the Panel on Confidentiality and Data Access (Duncan et al. 1993: 97-100) came to the conclusion that though researchers had a variety of informal means to influence how files were conceived, for example in the context of joint conferences involving researchers and civil servants from statistical offices, the discussion about institutional ways to involve data users had only barely begun. The Panel therefore recommended “to involve data users from outside the agency as statistical disclosure limitation techniques are developed and applied to data” (Duncan et al. 1993: 12).

Overview 2: Tasks of Research Community Institutions (Archives) in Data Acquisition

Country	Data Acquisition Concept	Influence on the Data Supplied	Participation in Ensuring Anonymity
France: LASMAS-IdL	Research interests of the CNRS laboratories and network	Bilateral negotiations with the statistical office and Ministries; comprehensive access to INSEE microdata; access to ministerial microdata a matter for negotiation	No institutional means
Great Britain: Data Archive, MIMAS	As comprehensive a supply of microdata to researchers as possible	Discussion with ONS and Ministries	For the 1981 Census: suggestions made by ESRC Working Group MIMAS: Secure data analysis in the network
Canada: DLI and affiliated universities (data libraries)	(Only the data libraries of subscribing universities have access to all DLI microdata)	Only Public Use Files. External Advisory Committee makes suggestions about data to be made accessible	No institutional means
The Netherlands: WSA	(WSA itself has no microdata; exercises function as broker)	WSA task: the expansion of the data now available in cooperation with users	WSA task: discussions with CBS about the „balance between confidentiality and user value“
Norway: NSD	If possible, universal supply to educational and research sector	Microdata archive is limited to surveys of individuals and households. NSD given data as needed. Access to microdata from administrative sources in consultation with national research community and data inspectorate	NSD receives microdata at the community level in anonymized form and engages in extensive additional measures on its own to ensure anonymity. Also responsible for linking survey data with other data sources
Austria: WISDOM	For its own research plans, all available microcensuses; other microdata upon request	WISDOM orders new microdata upon request by individual researcher or institute	No institutional means

Country	Data Acquisition Concept	Influence on the Data Supplied	Participation in Ensuring Anonymity
USA: ICPSR	Acquisition of all microdata needed by researchers (Public Use Files)	Recommended by the Panel on Confidentiality and Data Access	Institutionalization recommended by the Panel on Confidentiality and Data Access
Denmark: Danish Data Archive (DDA; Odense)	No dissemination of official microdata allowed		
Sweden: Swedish Social Science Data Service (SSD; Göteborg University)	No brokerage of official microdata established		

3.2 Financing

For researchers to have access to official microdata it is not only legal aspects that matter but financial ones as well. These affect not only the data preparation costs in the statistical offices but also the financing of institutions that provide data infrastructure.

A distinction can be drawn between the data infrastructure institutions of the research communities in Europe, where data services for research purposes are financed through public funds, and the situation in the USA, where the major social science service institution, the ICPSR data archive, is financed by subscribing and participating universities.⁶⁸ A financing form comparable to that in the USA was chosen by the Canadian Data Liberation Initiative (DLI) project, and aspects of an “institution-based funding“ can be found in the financial underpinnings of the Dutch Scientific Statistical Agency (WSA).

In terms of the principles used by the statistical offices for calculating costs, in general it is the additional costs for data preparation that form the basis for the fees. There is, however, a range from the provision of data at no cost (as in Great Britain and partly in the USA) to the calculation of proportionate costs for the data collection itself (as in Canada and the Netherlands). In addition, countries vary in terms of passing through the costs incurred in anonymizing Public and Scientific Use Files.

In individual countries, sometimes only after considerable effort, specific solutions have been

⁶⁸ Globally, the ICPSR is the largest social scientific data archive, and it provides a large variety of US and international data. The yearly subscription depends upon the size and type of university: those falling into Category A, “educational institutions with doctoral programs in the social sciences or related areas“, pay around \$11,000, while those in Category S, “Small colleges without doctoral programs in the social sciences or related areas and with enrollments of fewer than 2,500 students“, pay around \$3,500. See <http://www.icpsr.umich.edu/ICPSR/Membership/categories.html> for further details. All those who are affiliated with member universities, including students, can use all the services the ICPSR provides, which is to say they receive data and user support at no cost. Those belonging to subscriber institutions need to pay only half the cost for workshops and continuing education seminars (such as the ICPSR Summer

found, though they show large variation (see Overview 5). From the point of view of how the broadest possible use by the research community can be made of official microdata, it seems sensible to examine the various countries from the point of view of the individual researcher or research institution. One can distinguish between the following models:

1. The statistical office transfers the microdata to researchers at no cost.
2. The statistical office only calculates the additional (proportionate) costs incurred for preparing the data, whereby one can further distinguish between
 - a) an accounting with each individual user, or
 - b) financing through a data infrastructure institution of the research community
3. The statistical office calculates additional (proportionate) shares for the costs of the data collection itself, where one can also distinguish between:
 - a) an accounting with each individual user, or
 - b) financing through a data infrastructure institution of the research community

1) *The statistical office transfers the microdata to researchers at no cost.*

This model is primarily used in Great Britain, Norway, and, in essence, in the USA. The result is that where data is disseminated through data archives (as in, respectively, the Data Archive, the NSD, and the ICPSR), this can occur at a moderate cost that covers documentation and the data carriers. In the British case, this was based on an agreement reached in 1974 between the then statistical office OPCS and the research council ESRC, stating that “OPCS survey data are a valuable resource which should be made more widely available to the academic community. Deposit of such data in the Archive would increase the academic use by making them accessible and lessening the cost. The deposit would free OPCS from having to deal with repetitive applications“ (cited after Flaherty 1979: 59). This relationship between official statistics and researchers has subsequently frequently been expanded and reconfirmed (Church 1999; Sylvester 1996).⁶⁹

Since making microdata available is regarded as a public good in the USA, and regular publication as a duty, the statistical offices frequently provide Public Use Files for free on the Internet. For Public Use Files ordered directly from these offices, in general only minimal costs (such as for copying) or token fees need be paid. In the special case of providing

Program in Quantitative Methods).

⁶⁹ “Data Archive obtains data from ONS at no cost. This is an internal arrangement between two partners of government – not a market transaction. ONS is publicly funded, collects data from the public and it is government policy that as much use as possible should be made of this by researchers (...).“ (Answer from Alwyn Pritchard (ONS); 12 June 2000).

anonymized 1990 Census data to the ICPSR, the Bureau of the Census waived even calculating the fees, since, given the sheer size of the data files, it would have amounted to a considerable sum. In return, the ICPSR committed itself to providing comprehensive data documentation and user support (Austin/Rockwell 1997). Researchers from the ICPSR member universities can obtain the microdata at no cost; for all others, fees are charged.

2a) *The statistical office only calculates the additional (proportionate) costs incurred for preparing the data through an accounting with each individual user.*

In France,⁷⁰ Austria,⁷¹ Denmark,⁷² and Sweden⁷³ researchers must pay a data use fee per file, though in France this is only true of those researchers who are not associated with the CNRS.

2b) *The statistical office only calculates the (proportionate) costs incurred for preparing the data by financing through a data infrastructure institution of the research community (flat rate or inclusive fees)*

This model is used, in part, in Great Britain and France. In Great Britain the preparation of anonymized samples from the Census is financed centrally by the Economic and Social Research Council (ESRC) through the Office for National Statistics (Marsh/Teague 1992). Data can be used by the research community at no cost. In the same manner, the LASMAS-IdL in France receives microdata from the statistical office at a flat rate and can pass them on at no cost to researchers at the CNRS or associated institutions. Apparently, in the plans to create a microdata agency open to all researchers, the central financing model will be adopted; data for the use of individual researchers from this new agency will thereby become accessible at a lower cost than through the INSEE (see Appendix 1A).

3) *Calculation of proportionate data collection costs*

Type 3a, the accounting with each individual user, fits the Canadian case of researchers who

⁷⁰ Until recently, INSEE allowed researchers to pay reduced fees, but this has been rescinded as part of a general reduction of fees.

⁷¹ WISDOM calculates the cost of SPSS files authorized for dissemination by Statistics Austria at the same rates the statistical office charges for raw data.

⁷² Even though the research community in Denmark has access to extensive data resources with varied possibilities of data linkage and new improvements are to be noticed, according to Westergaard-Nielsen the most significant restrictions on data use are: most data are accessible at the statistical office exclusively and at considerable costs, only few researchers make use of the data, and the files are in hardly any cases available for comparative research at an international level. The provision of microdata as Scientific Use Files for use at universities or research institutions would be a better alternative (information obtained in the context of the KVI user survey).

⁷³ When asked what the ideal system of access to data would be, the Swedish researchers questioned in the context of the KVI user survey replied that the official data already collected for statistical purposes should be made available to the research community, whereby the fees charged should be limited to the costs arising for this specific provision of data.

must acquire Public Use Files directly from Statistics Canada when they are engaged in contract research that is externally financed and does not originate within a university. For university-based research in Canada and in the Netherlands, financing models are used which, in addition to the recouping of the data costs in the form of subscription fees or use fees, also contain elements of centralized financing (type 3b). With respect to the criterion of utilization-dependent cost calculation, the Dutch model would be closer to type 3a (accounting with each individual user), while the Canadian DLI model is closer to type 3b (financing through a data infrastructure institution of the research community).

Owing to changes in national budgeting priorities, the statistical office in the Netherlands has had to cover part of the data collection costs since the early 1980s by charging fees, now amounting to 1 million Guilders a year. This has had the effect, for example, of raising the fee for the Labour Force Survey from 6,000 Guilders in 1982 to 160,000 Guilders by 1988 (de Guchteneire/Timmermans 1990: 78) – a level too high for the research community to pay. One consequence has been a mediating initiative on the part of the Ministry of Education and Science: after a series of conversations, an agreement was reached between Statistics Netherlands and the Dutch research organization NWO to create the Scientific Statistical Agency (WSA) to act as the agency to approach for all distribution of microdata (Nobel 1995). The NWO covers the costs of the WSA. The estimated receipts of the CBS are financed through the central basic research support funds of the NWO. In return, the CBS makes at least 8 microdata files ready for the research community. For users engaged in academic research, data fees from 1,000 to 5,000 Guilders per file are levied, whereby acquisition of the entire available data as well as of older microdata can be obtained at reduced prices. Fees paid by institutes are largely used to recoup the costs incurred by the WSA.⁷⁴ Unlike the Canadian model, where all those who belong to the DLI subscribing universities can use Public Use Files, in the Netherlands it is not the university but a faculty that is the relevant organizational unit. Research institutes, even if they are linked to a faculty that is making use of data, must nevertheless sign their own contracts independent of the faculty (Scientific Statistical Agency 1996).⁷⁵

⁷⁴ Thus, for example, a research institute must pay 2,275 Euros for the 1998 Labour Force Survey, of which the CBS keeps 250 Euros for copying costs and the remainder is made available to the WSA (see appendix 4B). The costs double for non-academic institutions but are at different rates for specific data files.

⁷⁵ From the point of view of research the costs for the occasional use of data (by students, for example) are too high. Access to data is considered to be too bureaucratic. There are proposals that either the WSA and the Steinmetz Archive merge or that a government-financed data archive be created (replies by H. Ganzeboom

In Canada, after the mid-1980s, a change in the governing party led to a new fiscal policy that forced Statistics Canada to abandon its previous marginal cost calculations and low data dissemination fees, which made it cover part of the data collection costs through higher fees. In response, various universities, working with federal ministries, created a consortium to finance the Public Use Files from the Census, though this excluded all researchers not associated with consortium universities from access. Various initiatives to change such limitations led then in 1996 to the creation of the Data Liberation Initiative (DLI) project, as well as to a separate data agency within Statistics Canada (Albert et al. 1996; de Vries 1990; Statistics Canada 1997; Watkins/Boyko 1996). To resolve the problems of financing, the model of participatory subscription by universities was developed within the DLI. For university research itself, researchers and students at the subscriber universities had free access to all of Statistics Canada's Public Use Files.⁷⁶ Through the subscription fees, 70 percent of the central costs of the DLI are covered, with the remaining 30 percent coming from Statistics Canada and various Ministries.

From the point of view of an individual researcher interested in ready data access at moderate to no cost, the first three models (types 1 to 2b) are generally associated with minimal costs. When there is centralized financing,⁷⁷ the model also chosen in the pilot project in Germany, the British model (type 2b) seems the most attractive, and not merely for reasons of data costs. Compared with the Canadian and Dutch models (types 3a and 3b), it is more open to all interested researchers. Access to Canadian and Dutch microdata is only possible after the university has subscribed, or in the Dutch case after the faculty and the statistical office has reached an agreement.

and R. Muffels in the KVI user survey).

⁷⁶ The financing model employed by the DLI has some similarities to the member university subscription model used by the ICPSR.

⁷⁷ In this context it is worth noting that with the use of a central financing model, though it means an easier access to the data for the research community from the point of view of costs, it necessitates an accounting or coordinating office. Such an office needs to negotiate with the statistical offices about the extent of the data made available as well as set fiscal priorities under budgetary constraints.

Overview 3: The Financing and Costs of Disseminated Microdata

Country	Statistical Office: Basis for Cost Calculation or Fee Policy	Research Community: Financing Model
France	Calculation of additional costs incurred for anonymization, documentation, data copying, marketing; reduced prices for researchers abolished after fees generally lowered.	No financing model at present (other than in CNRS)
	For LASMAS-IdL/CNRS: Flat fee for all CNRS uses. Planned financing of preparation costs by INSEE through central research funds and proportionate fees from individual users.	LASMAS-IdL (financing from the CNRS budget): CNRS researchers have free access to the data; others pay fees for documentation and preparation
Great Britain	The Data Archive receives microdata at no cost. For the preparation of Census data, the ESRC as a central agency pays fees for the research community.	Data Archive: marginal costs for documentation, etc. MIMAS: free online utilization. Financing of the Data Archive and MIMAS through ESRC means and national infrastructure programs as well as money from sponsors
Canada	Since 1984 costs are covered and proportionate payments for data collection costs charged.	For DLI subscribers: for original university-based research, all Public Use Files are free; for smaller universities the yearly subscription costs \$3,000, for larger universities \$12,000. 70% of DLI financing comes from subscription; the rest from Statistics Canada and the government or Ministries
The Netherlands	Since the 1980s, fees contain a proportionate share for data collection costs. Money received by the CBS for microdata: 1 million Guilders	Contract between CBS and NWO to prepare at least 8 files each year. Creation of the WSA agency as a broker. Preparation fees contain a contribution to recoup WSA costs.
Norway	Statistics Norway prepares data for free	NSD provides users with the data for free. Fees for documentation and data carrier. NSD financed from the budget of the Norwegian research community.
Austria	Calculation of cumulative additional costs	No financing model. In ordering data through WISDOM, the same costs are incurred as ordering through Statistics Austria. WISDOM is financed by the Science Ministry.
USA	Fees usually cover only the cost of copying; in part freely available on the Internet	No financing model. Under a „Joint Statistical Agreement“: the ICPSR received the 1990 Census files at no cost. For researchers at ICPSR universities, microdata are free. ICPSR yearly contribution depend upon the size and type of university.
Denmark	All costs are covered	Researchers pay individually for 1) data preparation and 2) computer access
Sweden	Data which became part of „official statistics“ should be made available to the public at marginal cost	Researchers pay individually

3.3 Data Dissemination

There are three models for the dissemination of anonymized data to researchers (see Overview 4):

- 1) Transfer exclusively through the statistical office;
- 2) Transfer through the statistical office and alternately through an institution located in the research community;
- 3) Transfer exclusively through an institution located in the research community.

Whether official microdata can be disseminated not only through the statistical offices but also or exclusively through the data infrastructure institutions of the research community depends largely on the respective legal possibilities for doing so. If they exist, the infrastructure institutions, as in the case of Scientific Use Files, must be authorized by the statistical office through agreements that set forth the terms and conditions of data use and that contain obligations as to data protection.

1) Transfer exclusively through the statistical office

In the Netherlands, data is transferred exclusively by the CBS, though arranged through the WSA.⁷⁸ The Canadian model has data centrally held by the DLI, which is associated with Statistics Canada.⁷⁹ In France, data has thus far been provided directly from the INSEE.⁸⁰ However, the LASMAS-IdL provides extensive infrastructure services, from the acquisition of the microdata to the preparation for all CNRS institutes and affiliated researchers. In the context of the study contracted by the Ministry for Education, Research, and Technology [Ministre de l'Éducation Nationale, de la Recherche et de la Technologie] (Silberman 1999), the creation of a general social science institution for official microdata, open to all researchers, is currently under discussion (see appendix 1A and 1B). With very few exceptions, the Danish statistical office does not provide microdata to researchers. Microdata that are made available can only be used in the data analysis centers of Danmarks Statistik itself. The Swedish statistical office rejects the idea of providing data through the social

⁷⁸ Data documentation and user support are regarded as unsatisfactory, according to the researchers H. Ganzeboom and R. Muffels, questioned in the context of the KIV user survey.

⁷⁹ The creation of the agency within the statistical office has to do with the fact that Statistics Canada alone is legally empowered to disseminate Public Use Microdata Files. It is also noteworthy that no national data institution exists within the research community, and until the creation of the DLI project in Statistics Canada, there was not even a central office responsible for the dissemination of Public Use Microdata Files (see de Vries 1990). Thus in an institutional sense, the DLI project resolved an organizational problem that had existed at Statistics Canada regarding access to Public Use Microdata Files.

⁸⁰ The national social science archive CIDSP/BDSP concerns itself only with research surveys.

science data archive at the University of Göteborg, giving concerns about controlling access to the data and loss of income as reasons.⁸¹

2) *Transfer through the statistical office and alternately through an institution located in the research community*

Public Use Files in the USA are made available both through the statistical agencies and from data archives like the ICPSR. Data from the National Longitudinal Study, which like anonymized German microdata are only accessible under contractual term, use limitations and data protection obligations (e.g., falling under the rubric of „restricted data access“), are contractually or cooperatively provided from the Bureau of Labor Statistics through the Center for Human Resource Research at Ohio State University (see appendix 7A). In Austria, the WISDOM has no greater data rights than any others who acquire data, but can pass on copies or excerpts of its microcensus data to individual researchers with the approval of Statistics Austria.

3) *Transfer exclusively through an institution located in the research community*

Particularly in Great Britain and Norway, the research community institutions that provide data services to users (such as the UK Data Archive, the MIMAS, or the NSD) exercise very comprehensive trusteeship functions (see Overview 4). In both countries, the archive occupies a key role between the statistical offices, as the data producers, and the researchers, as the data “consumers” or users. The statistical offices regard this as beneficial (see Church 1999) since the amount of work they need to do for each individual researcher is considerably reduced, at least compared with providing data directly from the statistical office.

In general one can conclude that in most countries the infrastructural institutions of the research community play important roles in the dissemination of official microdata. In Norway the statistical office even prefers to pass on data via the data archive, and in the United Kingdom the statistical office does so exclusively through the UK Data Archive. In Canada and the Netherlands, data dissemination is exclusively through the statistical offices, though in Canada a separate office was created within Statistics Canada in the context of the DLI project⁸² and in the Netherlands the WSA functions as a broker between the CBS and the

⁸¹ From the point of view of researchers a data file that was very expensive to produce and was financed through public funds should not be used by one single project only; it should be made available to other projects, too (information supplied by N.A. Klevmarken in the context of the KVI user survey).

⁸² The grounds for this are on the one hand that exclusive Crown Copyright for Public Use Microdata Files is in the hands of Statistics Canada, and on the other the lack of a national data archive within the research

research community.

For a researcher interested in official microdata, it is not only important to know which office or agency to turn to but also in which form the data may be obtained. Here one needs to distinguish between long-term archiving and current preparation of data. In the case of the former, it is essential that data and documentation are extant in a form that is independent of any software system. For the latter, it is important that data and documentation are integrated and available in a documented file format. The statistical offices usually have too few resources to perform both functions. In some countries microdata is provided largely in the form of raw data without further machine-readable file documentation.⁸³ In other countries, such as in France, file formats and degrees of integration vary. Data infrastructure institutions within the research community, on the other hand, are responsible not only for securing longer-term availability but also for current preparation in user-friendly file formats. As a result they assume the responsibility for controlling and preparing the data, as well as for creating documentation in (an increasingly internationally) standardized form,⁸⁴ and make data and documentation in current program formats available to the end users. If one considers the amount of time and personnel needs associated with data preparation, this work relieves individual researchers not only of the burden of processing the data themselves but it also avoids duplication of effort. Not the least important service, in the context of data preparation, is provided by data consistency checking, augmenting the plausibility controls provided by the data producers: "Since researchers often use the data in ways that are different from the original data collection objective, the data are subjected to a variety of edits and checks that are not likely to be made by regular processing" (McGuckin 1993: 86). That researchers in the USA acquire Public Use Files more frequently from the ICPSR than from the statistical offices is due, in the estimation of the ICPSR itself, to the careful data preparation and documentation this data archive undertakes (on this, see Erik Austin, appendix 7B).

community.

⁸³ The U.S. Census Bureau has, however, launched a demanding program to try make user information and metadata available: "There is no question that Government agencies should no longer provide just raw data, they should provide value-added information. (...) The Census Bureau, with its massive stores of data, technology, and user service expertise, is in a most advantageous position to meet the American public's emerging information needs – to move beyond mere data dissemination to knowledge dissemination" (abstract of Wallace et al. 1999).

⁸⁴ See the Norwegian Social Science Data Services 1999.

Overview 4: The Organization of the Transfer of Public Use and Scientific Use Files to the Research Community

Country	Data Transfer Offices	Data Formats	Terms of Use Contracts
France	INSEE or Ministries. For CNRS researchers: LASMAS-IdL. Planned: Creation of a universal microdata agency in the research community	INSEE: raw data and all current file formats; LASMAS-IdL: ditto	No contract required for Public Use Files. For Scientific Use Files: INSEE or Ministries. For CNRS researchers: LASMAS-IdL
Great Britain	Data Archive (transfer on data carriers), MIMAS (online access to Census data)	Data Archive, MIMAS: SAS and SPSS files, data bank files	Through Data Archive or MIMAS (no data provided directly through the statistical office)
Canada	For DLI universities: FTP access to the DLI server; otherwise directly from Statistics Canada	DLI: raw data, SAS and SPSS Setups	DLI: through local DLI contact; otherwise through Statistics Canada
The Netherlands	CBS (via WSA)	Raw data, SPSS Files	Through CBS
Norway	NSD (usually; otherwise SSB)	NSD: SPSS, SAS and NSDstat as well as data bank files	Completion preferred through NSD (otherwise through SSB)
Austria	Statistics Austria WISDOM (only with authorization)	Statistics Austria: raw data. WISDOM: raw data, SPSS files	Through Statistics Austria
USA	For Public Use Files: statistical offices and archives (e.g., ICPSR). For Scientific Use Files: statistical offices and archives in part or research institutions working for the statistical offices	Statistical offices: raw data. ICPSR and other Archives: raw data, SAS and SPSS Setups	Required only for Scientific Use Files (through statistical offices; partly through research community data archives)
Denmark	Only access on site in Statistics Denmark.		Statistics Denmark
Sweden	Statistics Sweden. No Public Use Microdata Files		Statistics Sweden

3.4 Expansion of Analytic Potential

Beyond the above-mentioned work in data preparation and documentation, additional services are provided by the research community's institutions in the context of supplying data which expand the possibilities for analysis (see Overview 5). With respect to the UK Data Archive, Denise Lievesley described these services as follows: "We work with data providers to assist them to produce full and well structured documentation. In addition we ensure accessibility by using demand led distribution systems, and we provide value added services by integrating different datasets, adding contextual information, reformatting for data delivery, extracting subsets of data and documentation, and by enabling users to visualize, browse and select

data“ (1998: 311).⁸⁵

As one example, official classification systems and typologies are often not directly usable in social and economic research, but do provide a basis for research-specific classifications or scales. It is standard practice for most social science data files to include such research-oriented classifications, unlike in official data. These tools can be disseminated either within the research networks⁸⁶ or through the archive itself. Research community institutions like the UK Data Archive or the MIMAS (see Middleton 1995) can expand the potential use of the data considerably by recoding the microdata and thereby relieve the user of this routine work.⁸⁷

Research needs also exist in terms of the diachronic comparison of data. Official microdata do display considerable advantages for the investigation of historical trends, but because the work of statistical offices is largely oriented towards the generating of current statistics, little is done to address the problems of comparing data over time. The development or preparation of comparable (or trend) datasets against this background is an important task for the institutions that provide services to researchers. The ICPSR, for example, provides Current Population Survey and historical Census data in a form standardized through research projects. Another example is provided by bringing together the individual waves of the rotating panel samples, indicating how individual research projects can increase the potential use of official data and achieve synergistic effects. Though longitudinal investigations like the Current Population Survey and the Austrian microcensus are conceptualized as rotating panel samples, this characteristic is only partly exploited by the statistical offices themselves (see Madrian/Lefgren 1999, and in contrast, Tate 1999).

The cooperation between the Norwegian NSD data archive, the statistical office, the social security bodies, and the Health Ministry in linking survey data with administrative sources (Kvalheim 1999) is worth particularly emphasizing, because it makes evident that official statistics as well as the research community can profit greatly thereby. With such intensive

⁸⁵ In the KVI user survey, C. Mills and H. Sutherland consider the data documentation and user support provided by the UK Data Archive to be very good. According to H. Sutherland, this should be improved in the future by providing more support with respect to "expert" questions as well as a more detailed data documentation and additional training regarding the utilization of complex data (such as panels).

⁸⁶ See, for example, the website of H. Ganzeboom on occupational classification at <http://www.fss.uu.nl/soc/hg/ismf/index.htm>.

⁸⁷ That the data dissemination offices within the Canadian and Dutch statistical offices provide no additional information, such as research-specific variables which could count as "added value", probably is related to the neutrality precept followed by official statistics.

forms of cooperation, an enlargement of the data utilization potential can be realized.⁸⁸

Overview 5: Expansion of Analytic Potential

Country	„added value“
France	LASMAS-IdL: construction of quasi-longitudinal files (trend data) from cross-sectional data
Great Britain	ONS: cooperation with researchers in the creation of a longitudinal file of the Labour Force Survey (Tate 1999) Data Archive, MIMAS: the passing on of research-specific variables
Canada	-
The Netherlands	-
Norway	Cooperation between SSB and NSD in keeping health register; the linking of survey with register and administrative data; keeping a comprehensive longitudinal data bank linked to labor market and social security data (KIRUT event data bank)
Austria	WISDOM: user support in linking rotating panel waves of the microcensus
USA	ICPSR: preparation of standardized, comparable trend datasets (in the Current Population Survey Uniform March Files 1964-1988, Uniform October Files 1968-1990) IPUMS: integrated public microdata series (historical Census data)
Denmark	-
Sweden	-

3.5 Consultation Services Provided

The appropriate operationalization and framing of questions, relative to the data basis, means there is a need for consultation and advice, and it is a need that affects the entire research process - from the hunt for data to problems of methodology.

All statistical offices have persons available to answer questions from researchers,⁸⁹ but only the Dutch statistical office provides consultation services beyond initial inquiries.⁹⁰ As Overview 6 indicates, the range of consultative services in the data archives is by contrast comprehensive and oriented toward user needs. Worth noting particularly is a new “Training & Dissemination Project“ which in an innovative fashion links information about official microdata with methodological issues and statistical modelling (Wiggins et al. 2000).

Since the beginning of the DLI project in Canada, the local contacts (that is, in the university data libraries) have played an important role as the intermediaries between the DLI Unit

⁸⁸ The NSD’s special competence regarding user support and data documentation is substantiated by particularly positive judgments coming from the researchers who were questioned (information supplied by Professor Dr. J.E. Kolberg and Professor Dr. A.E. Risa in the context of the KVI user survey).

⁸⁹ As Erik Austin (see appendix 7B) notes, the user services provided by the statistical offices in the USA is inadequate, viz. “the user services they offer are usually nonexistent”.

⁹⁰ Included among the tasks of CBS statisticians is also the review of research reports (including conference and lecture manuscripts) prepared on the basis of research conducted with the microdata. Following the contract agreed upon between CBS and the research institution, data users must provide the CBS with a copy of the research report prior to publication to enable the CBS to review the content for its maintenance of statistical confidentiality. The CBS has 15 working days in which to do so, and must within this time communicate suggestions for revision, if they are deemed necessary.

within Statistics Canada and the researcher.⁹¹ The increase in the utilization of microdata has also been accompanied by an increased demand for consultation (Strike 1998; Fichter 1998). To meet such demands, particularly in questions about methods, continuing education for the personnel at the local DLI contact points has been undertaken and informational networks created.⁹²

In their preparation of data documentation, the taking on of consultation tasks as well as other service provision,⁹³ the data archives augment as well as relieve the statistical offices not only of routine tasks but also efficiently meet research-specific needs. Denise Lievesley (1998: 311) has argued that “Data dissemination centres such as ours can act as a buffer between scientific users and producers. One might think this is a barrier between the two groups. I prefer to think it as an intelligent filter which is desirable because many queries and problems are unrelated to the data and because supporting users is time-consuming and requires an understanding of their needs“. Scientific data archives can exercise this supportive function particularly well when they possess not only expertise about official microdata but themselves also conduct limited research, thereby building a better basis for evaluating the microdata from a user’s perspective.

Workshops and user conferences which include those who themselves are data producers serve not only as venues to exchange information and experience among data users, but also serve as a means for professional statisticians to communicate technical information. They provide a feedback mechanism for the statistical offices that, through the discussions, are able to hear useful suggestions in how they can improve their surveys and investigations.⁹⁴

⁹¹ This role corresponds with the so-called “Official Representatives” of the ICPSR who are not only the local user contacts with the ICPSR, but who also represent the university to the ICPSR.

⁹² The SHERLOCK analysis system was developed by the Quebec university libraries (see <http://sherlock.crepuq.qc.ca/empdescr.html>) to make the microdata made available by Statistics Canada through DLI more open and usable. At the University of Alberta’s data library two workshop modules for working with DLI data are also available (see <http://datalib.library.ualberta.ca/accoleds/workshops/>).

⁹³ To these service aspects one can also add special evaluations. Archives like WISDOM also conduct special surveys under contract for researchers who lack the necessary resources (computer, know how) to process large quantities of official microdata.

⁹⁴ Judging by the answers provided by the statistical offices (see appendix), there could be further improvements in the feedback mechanisms from users to data producers regarding data analysis. Most effective for the statistical offices, apparently, are direct cooperation in the research itself as well as the organized feedback provided by workshops.

Overview 6: Consultation Services Provided by Statistical Offices and Data Archives

Country	Statistical Office	Data Archive
France	Information provided largely with respect to technical questions	LASMAS-IdL: user support, workshops
Great Britain	Participation in user meetings organized by the Data Archive and MIMAS	Data Archive, MIMAS: user support, mailing list, workshops and user meetings; data-oriented methods training (Wiggins et al. 2000)
Canada	Central information in the DLI: Queries sent to STC specialized department and information disseminated over the DLI network. Regional workshops held in conjunction with the local DLI contacts	Local DLI contact: user support, preparation of sample evaluations; knowledge exchange through the DLI network
The Netherlands	Individual user consultation by data and methods experts; participation at WSA user meetings	WSA: user meetings, mailing list
Norway	Information provided by specialized departments	NSD: technical support and methods consultation; clarification over data protection questions
Austria	Information provided by specialized departments	WISDOM: content and methods consultation; support in the linking of microcensus panel waves; evaluation service
USA	Information provided by email contact addresses; partial Internet online data analysis tools	ICPSR: evaluation support, workshops. Center for Human Resource Research (contracted by the Bureau of Labor Statistics): user services for the National Longitudinal Survey
Denmark	User services	
Sweden	User services	

3.6 Assessment of Data Access by Researchers

The various means by which official microdata are disseminated to the research community, at least from the point of view of data users, can be readily summarized. The KVI user survey focused on data offered, data access, costs, and user support, but since few answers were given regarding cooperation between official statistics and the research community, this particular topic is omitted here (although it is discussed elsewhere in our report). The most significant findings are reproduced in Overview 7 below. In interpreting these findings, one should of course note that the answers provided by the various specialist respondents in each country reflect their own expertise and experience; some answers are therefore incomplete and answers vary in their comprehensiveness. Bearing these limitations in mind, the following picture emerges:

Data Supply

Researchers were very positive in their assessments not only of the possibilities for analyzing microdata from administrative sources, but also about the usable data on offer. For data made available through the statistical office, significant restrictions on analysis were mentioned only in the Canadian and Dutch cases. There was general support for making Scientific Use Files available to a broad spectrum of research projects and agendas.

Data Access

Where data access is organized through an institution within the research community, as in Great Britain and Norway, access is judged as good to very good. Where data access is exclusively provided through the statistical office, including the requirement that the data could only be worked on in the statistical office itself, respondents usually made the connection between access to data and restrictions on its use.

Costs

It should be acknowledged that the use of microdata stemming from special evaluations, as in Canada, or the creation of project-specific files, as in Denmark and Sweden, is associated with high data costs. Where data is supplied by the statistical office and users receive individual invoices, fees can be quite high, but when data is supplied through an institution within the research community, individual users need bear only low to moderate costs.

User Support

One finds a similar pattern in the assessment of user support as one finds in the assessment of the organization of data dissemination. But most researchers judged the documentation supplied by the statistical offices as well as the consultative services provided by their specialized offices as rudimentary. However, research-specific consultation and data documentation supplied by the UK Data Archive as well as by its Norwegian counterpart were judged good to very good.

Overview 7: Researcher Assessment of Data Access, Costs, and User Support

Country	Data Access	Costs	User Support
<i>1. Data provided by the statistical office</i>			
Canada	<ul style="list-style-type: none"> • Good data supply. Restrictions on analysis of Public Use Files. Improvements through the DLI and data analysis centers. • Disadvantages of special evaluations: missing replication possibilities. 	<ul style="list-style-type: none"> • High costs for Public Use Files (for those outside the DLI) and for special evaluations. 	<ul style="list-style-type: none"> • DLI: no user support
The Netherlands	<ul style="list-style-type: none"> • Restrictive data access. • Serious limitations on analysis owing to anonymization process. • Improvements brought about through creation of WSA. • Disadvantages: bureaucratic, time and effort. • Suggestions: independent data archive 	<ul style="list-style-type: none"> • Data fees are too high when there is only occasional use. • Expensive financing (through NWO) for access to company data in the CBS data analysis center. 	<ul style="list-style-type: none"> • Insufficient data documentation and missing user support (the reason: the WSA has no data and conducts no analyses)
Denmark	<ul style="list-style-type: none"> • Comprehensive analysis possibilities. • As a rule, data use only in the statistical office. • Advantages: data linkage, good work conditions. • Disadvantages: little use of data, no standard files. 	<ul style="list-style-type: none"> • Moderate fees for computer use in the data analysis centers. • Overall relatively high costs (for data production combined with computer use fees) 	<ul style="list-style-type: none"> • User support provided by professional statisticians in the data analysis centers.
Sweden	<ul style="list-style-type: none"> • Broad data on offer and linkage possibilities. • Data access situation satisfactory. • Disadvantage: no standard files. 	<ul style="list-style-type: none"> • For many researchers (particularly students), the costs are too high. • Marginal cost calculation principle is not always adhered to. 	<ul style="list-style-type: none"> • Minimal data documentation provided by the SCB.
France	<ul style="list-style-type: none"> • Comprehensive data offered. • Regulated access to INSEE data, in part difficult access to data from Ministries. 	<ul style="list-style-type: none"> • Moderate costs through LASMAS or CNRS-associated institutes. • For professorial research outside CNRS the costs are too high. 	<ul style="list-style-type: none"> • INSEE: minimal user support and documentation. • LASMAS: comprehensive user support and documentation.
<i>2. Data provided through statistical offices as well as through institutions within the research community</i>			
U.S.A.	[Suggested increase in the dissemination of Scientific Use Files; David 1998]	<ul style="list-style-type: none"> • Analysis centers: much time needed for task completion. 	
<i>3. Data provided through an institution with the research community</i>			
Great Britain	<ul style="list-style-type: none"> • Comprehensive data on offer • Good data access through the Data Archive. • Would be helpful to have Public Use Files available for teaching purposes. 	<ul style="list-style-type: none"> • Data Archive only charges minimal fees for Documentation 	<ul style="list-style-type: none"> • Data Archive: good basic documentation and user support. • For specialized questions, a lack of detailed information and support.
Norway	<ul style="list-style-type: none"> • Comprehensive data access. • Division of labor between the statistical office and the NSD for researchers very satisfactory. 	<ul style="list-style-type: none"> • Minimal or no fees. 	<ul style="list-style-type: none"> • NSD: very good documentation and user support.

4 Researcher Access to Social Security Microdata

For economic and sociopolitical research, the data from social insurance agencies and other systems of social protection are of extraordinarily large significance both for methodological reasons⁹⁵ and in terms of their thematic foci (Schmähl/Fachinger 1994). In Germany, with the exception of the IAB employment sample available through the Central Archive (ZA) in Cologne, the access to such data is not regulated. An overview whether or to what extent social welfare and social security microdata is accessible to researchers in other countries can thus provide important stimuli. Given the complexity of what are, in international comparison, quite differently organized welfare systems, here we can only provide a preliminary overview. A detailed depiction of access possibilities and access forms that takes institutional differences explicitly into account was not possible within the time frame for this report. The various legal frameworks could also not be worked up.⁹⁶

Our survey was limited to a brief questionnaire asking about the possibility or form of access to microdata covering pensions, unemployment compensation, and social assistance.⁹⁷ In addition to the countries surveyed previously in this report, Finland and Switzerland were also included. Scandinavia seemed to us to be particularly important to include since the integration of administrative sources is quite advanced and data protection comparatively stringent. There are also varied experiences in the dissemination of such data to researchers in Scandinavia. Switzerland was included in order to provide complete coverage of German-speaking countries.

Summary of the Conclusions

With the exception of Great Britain⁹⁸ (but not of Northern Ireland), all the aforementioned countries permit researchers access to social security microdata, though access in Switzerland is still in the trial phase. In Denmark no direct access exists to such microdata, yet there is access to research data banks of the statistical office in which social security microdata are

⁹⁵ On the connection between life-cycle and multilevel analysis, see Mayer/Huinink 1994.

⁹⁶ In this context, see Eurostat 1997.

⁹⁷ The authorities responsible for the various social services were contacted, not their supervisory bodies. This was done to minimize response time and effort.

⁹⁸ Bona fide researchers in Great Britain do have an alternative source available in the regularly-conducted Family Resources Survey and its follow-up surveys, as well as the linkages with administrative sources (see Department of Social Security 1998). For unemployment data and services provided to the unemployed, the statistical office also has available a cohort data bank (JUVOS). See http://www.statistics.gov.uk/themes/labour_market/surveys/juvos.asp.

integrated.⁹⁹

With the exception of Northern Ireland, researchers have cross-sectional as well as longitudinal data at their disposal.¹⁰⁰ In the majority of the countries examined, it is permissible to link survey data with administrative sources for both cross-sectional as well as longitudinal data,¹⁰¹ so that a considerable information potential is available. Access and forms of data supply vary not only between countries but also by modality. Frequently, the data is prepared by those agencies that are themselves responsible for social security and social assistance, though often as Scientific Use Files that can only be used for a limited time and under contractual rules. Standard files tend to be the exception, data extracts made for specific research projects the rule. Sensitive data are usually only available in restricted areas of the data provider agency and only to qualified researchers.

Research is supported particularly well at the moment in Scandinavia, Great Britain, France, and the USA. Throughout Scandinavia, researchers have thematically varied event data banks at their disposal, whose analysis potential can be expanded, when justified by the research needs with corresponding linkages to other microdata sources.¹⁰² In Great Britain, with the ONS Longitudinal File¹⁰³ and the JUVOS event data bank,¹⁰⁴ researchers have very interesting data to work with. In the USA, the Social Security Administration has created two multidisciplinary research centers¹⁰⁵ (on a trial basis) for analyzing the problems of social security, particularly pensions, and these centers can draw on the Social Security microdata. The following synopsis summarizes in a structured form what the most important conclusions are that can be reached from our questionnaire.¹⁰⁶ In interpreting it, it is important to remember that the questionnaire did not *explicitly* address the different schedules. Instead, the institutional components are only indirectly visible in the details provided about the program

⁹⁹ For the Fertility Database, see Danmarks Statistik 1999; for the Integrated Database for Labour Market Research, see Danmarks Statistik 1991.

¹⁰⁰ In France access to welfare information is limited to cross-sectional data.

¹⁰¹ In Norway, it is generally not permitted to link welfare and unemployment data; in Canada, the connections are limited to cross-sectional data; in the Netherlands, though linkage is permissible, it is not done.

¹⁰² For example, the KIRUT data bank in Norway, and the administrative sources that are being assembled on health research there (Kvalheim 1999), the consolidated Labour Statistics Register in Finland (The Social Security Institution of Finland/Kela: questionnaire answer) and the Finnish Longitudinal Census Data File (Statistics Finland: n.d.), and the Fertility Database (Danmarks Statistik 1999) and Integrated Database for Labour Market Research in Denmark (Danmarks Statistik 1991).

¹⁰³ See <http://www.cls.ioe.ac.uk/Ls/lshomepage.html> .

¹⁰⁴ See http://www.statistics.gov.uk/themes/labour_market/surveys/juvos.asp .

¹⁰⁵ The Retirement Research Consortium at <http://www.ssa.gov/policy/research/RRC/index.html> .

¹⁰⁶ A brief overview of the questionnaire and the responses received may be found in the concluding remarks at

responsibilities of the agencies addressed.¹⁰⁷ It would be desirable to gather information that differentiated explicitly by type of scheme but without the support of German and international umbrella organizations, it would be virtually impossible to implement.

the end of the synopsis.

¹⁰⁷ The answers provided by umbrella organizations like the Swiss Social Security Ministry, or by supervisory bodies such as the Dutch Social and Labor Ministry (SZW) tended to be superficial.

Overview 8 Access to microdata from social security bodies in selected countries: synoptic summary of a questionnaire-based inquiry

Country	Agency inquired of	Program responsibility of agency	Cross-sectional data		Longitudinal data		Answers to open questions; Notes indicated by brackets
			Access for bona-fide research	Linking of admin. records to survey data: legal provisions	Access for bona fide research	Linking of admin. records to survey data: legal provisions	
Austria	Hauptverband der österreichischen Sozialversicherungsträger	Pension and sickness insurance	Yes	Linking to files of the Ministry for Social Security and Generation Issues (BMSG) permitted.	Yes	Linking to files of the Ministry for Social Security and Generation Issues (BMSG) permitted.	Access to microdata only with the permission of the Ministry for Social Security and Generation Issues [Bundesministerium für soziale Sicherheit und Generationen; BMSG]
	Arbeitsmarktservice	Unemployment insurance					(no reply so far)
	Bundesministerium für Arbeit, Gesundheit und Soziales	Social assistance					No data available at national level.
Denmark	Arbejdsmarkedets Tillaegspension (ATP)	Supplementary pensions (earnings-related pensions)	No		No		Access for bona-fide research not under discussion. The agency carries out its own research.
	Direktoratet for Arbejdsloshedsforsikring	Unemployment insurance	No		No		Opening access for bona-fide research not under discussion. The agency carries out its own research
	Statistics Denmark (Publication: 'The Fertility Database' Copenhagen, 1998)	All social protection benefits (incl. event history data)	Access for scientific purposes possible through special database of Statistics Denmark ('The Statistical Register for Fertility Research'). Upon request, additional data can be provided by Statistics Denmark's Integrated Statistical System on Persons.				Special arrangements for researchers possible under certain conditions for research on the premises of Statistics Denmark. Provision of individual tables upon request.
Finland	The Social Insurance Institution of Finland (KELA)	All social insurance benefits (including social assistance).	Yes	Linking not permitted.	Yes	Linking permitted.	Scientific Use Files available directly from KELA or indirectly from Statistics Finland. Research with confidential data possible under certain conditions at the premises of KELA. Statistics Finland has established various longitudinal databases and conducts its own research.

Country	Agency inquired of	Program responsibility of agency	Cross-sectional data		Longitudinal data		Answers to open questions; Notes indicated by brackets
			Access for bona-fide research	Linking of admin. records to survey data: legal provisions	Access for bona fide research	Linking of admin. records to survey data: legal provisions	
France	Caisse Nationale d'Assurance Vieillesse (CNAV)	Old age pension insurance	Yes	Linking to survey data of the DREES/INSEE permitted. [Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques; DREES]		Yes	Samples of data transferred to Statistics France (INSEE) upon permission by the CNIL. Linking of data organized in such a way that confidentiality is safeguarded. CNAV conducts its own research.
	Caisse Nationale des Allocations Familiales (CNAF)	Social assistance	No	Linking to survey data of the Ministry of Housing permitted.	No	No	Passing on of data to the Ministry of Social Affairs with certain restrictions (experimental) CNAF conducts its own research
	Union Nationale Interprof. pour l'Emploi dans l'Industrie et le Commerce (UNEDIC)	Unemployment benefits					(No response so far.)
Great Britain	Department of Social Security	Pension insurance and social assistance	No		No		Opening of access to external researchers still under discussion. Ministry currently offers comprehensive analysis services and tabular data upon request, including small area data. Ministry conducts its own research and carries out its own statistical surveys.
	Department of Education and Employment	Unemployment insurance					(No response so far.) Ministry conducts its own research and carries out its own statistical surveys.
Northern Ireland	Department of Social Development	Pension and unemployment insurance	Yes; Scientific Use Files in certain	Passing on of data to ONS and ministries for linking to survey data	No		Provision of data to external researchers only where appropriate confidentiality agreement has been signed. No information av. on own research activities

Country	Agency inquired of	Program responsibility of agency	Cross-sectional data		Longitudinal data		Answers to open questions; Notes indicated by brackets
			Access for bona-fide research	Linking of admin. records to survey data: legal provisions	Access for bona fide research	Linking of admin. records to survey data: legal provisions	
The Netherlands	Ministerie van Social Zaken en Werkgelegenheid	Pension, sickness and unemployment insurance (supervision); Social assist. (administrative responsibility).	Yes; available either from Ministry or from Statistics Netherlands.	Linking to survey data permitted, but not used so far.	Yes; anonym. data via WSA from Statist. Office; access to confid. data via on site research at Statistics Netherlands.	Linking to survey data permitted.	(Information covers all social security schemes under supervision of the Ministry.) Access to confidential data possible under certain conditions and with permission of Ministry. Debate on provision of administrative records of social security schemes for research purposes still continuing.
Norway	Royal Ministry of Labour- and Government Administration	Unemployment insurance	Yes; Files available from stat. office and from social science information service NSD.	Linking to survey data not permitted.	Yes, available from Statistics Norway or NSD.	Linking to survey data permitted.	Provision of longitudinal files still experimental. The Ragnar Frisch Centre for Economic Research has special databases on unemployment (including event history data) based on linked data from surveys and administrative sources. Cf. http://www.frisch.uio.no
	The National Insurance Administration (NIA)	Pension and sickness insurance	Yes; files available from Statistics Norway or from NSD; in special cases, files with selected variable sets directly available from NIA	Linking to survey data permitted	Yes; files available from Statistics Norway or NSD; in special cases, files with selected variables available from NIA.	Linking to survey data permitted.	The KIRUT database maintained by NSD comprises event history data from all social security branches. NSD health database with linked data from surveys and administrative data in stage of formation. (Cf. Kvalheim 1999). No information provided regarding research activities of Statistics Norway.
	Statistics Norway (SSB)	Social assistance	No.		Yes; files available directly from Statistics Norway or from NSD.	No.	

Country	Agency inquired of	Program responsibility of agency	Cross-sectional data		Longitudinal data		Answers to open questions; Notes indicated by brackets
			Access for bona-fide research	Linking of admin. records to survey data: legal provisions	Access for bona fide research	Linking of admin. records to survey data: legal provisions	
Sweden	Riksförsäkringsverket	Pension and sickness insurance	Yes; files available directly from Statistics Sweden.	Only Statistics Sweden is permitted to link administrative records to survey data.	Yes; anonymized data files available directly from Statistics Sweden.	Only Statistics Sweden is permitted to link administrative records to survey data.	In special cases – and only with the permission of all persons affected (informed consent) – provision of confidential data for research purposes possible (cross-sectional data as well as longitudinal data). Agency does not conduct its own research.
	Arbetsmarknadsstyrelsen	Unemployment insurance	Yes; Scientific Use Files with selected variable sets available directly from administering agency.	Only Statistics Sweden is permitted to link administrative records to survey data.	Yes; Scientific Use Files with selected variable sets available directly from administering agency.	Only Statistics Sweden is permitted to link administrative records to survey data.	In special cases and under certain restrictions provision of confidential data possible for research purposes. No information supplied on its own research activities.
	Socialstyrelsen	Social assistance	Yes; Scientific Use Files with selected variable sets from administering agency or Statistics Sweden.	Only Statistics Sweden is permitted to link administrative records to survey data.	Yes; Scientific Use Files with selected variable sets. from administering agency or Statistics Sweden.	Only Statistics Sweden is permitted to link administrative records to survey data.	Special permission by the Board of Health and Welfare required. No access at all possible to individual records of the Social Assistance Scheme. Agency conducts its own research.
Switzerland	Bundesamt f. Sozialversicherung (BSV)	Pension, sickness and unemployment insurance	Yes, through the Federal Statistical Office.	Only the Federal Statistical Office is permitted to link administrative records to survey data.	Yes; available from the Federal Statistical Office.	Only the Federal Statistical Office is permitted to link administrative records to survey data.	Access is in stage of formation and will vary between the different social security schemes. Currently the development of a microsimulation model for the old age pension system and the widows' pensions as well as the linking of administrative records to the labor force survey of the Swiss Federal Statistical Office are in the planning stage. BSV conducts its own research.

Country	Agency inquired of	Program responsibility of agency	Cross-sectional data		Longitudinal data		Answers to open questions; Notes indicated by brackets
			Access for bona-fide research	Linking of admin. records to survey data: legal provisions	Access for bona fide research	Linking of admin. records to survey data: legal provisions	
Canada	Human Resources Development Canada (response by Statistics Canada)	Pension and unemployment insurance	Yes (Scientific Use Files); research with personal data possible on the premises of Statistics Canada.	No linking to survey data permitted.	Yes (as Scientific Use Files); access to confid. data under certain conditions possible on the premises of Statistics Canada.	No linking to survey data permitted.	Access to longitudinal data files in experimental stage and closely related to the establishment of dedicated research data centers.
USA	Social Security Administration (SSA)	Pension insurance	Yes; scientific use files from the Michigan Retirement Research Center hosted by the Institute of Survey Research.	Only the Census Bureau is permitted to link under certain conditions administrative records to survey data.	Yes, on the premises of the agency (selection of variables).	Only the Census Bureau is permitted to link under certain conditions administrative records to survey data.	Access under certain conditions: <ul style="list-style-type: none"> - Research must be in line with the tasks of the SSA; - use of linked data only possible as 'special sworn employee' - Publication of results only if data protection rules are not violated. SSA finances two research centers dealing with retirement (Retirement Research Center at Boston and at Michigan). (http://www.mrrc.isr.umich.edu/intro.html)

Notes on Process and Response to the Questionnaires

For ease of analysis, standardized questions with prepared response categories were mainly used. In all, three basic situations of data access were elicited, each with their specific component parts:

- Is there access at all? If not, has such access been discussed lately and what was the (significant) outcome?
- Is there only trial access? If so, what are the basic arrangements of this trial?
- If access existed, further questions addressed the form that access and dissemination took, the possibilities or permission of linkage with survey data, and the key conditions or limitations for use (open question). In this, questions distinguished between cross-sectional and longitudinal data.

Independently of the data access questions, an additional question was asked whether the respective institution conducted its own research.

The questionnaires were usually in English, though a French version was prepared for France and a German version for Austria and Switzerland, and were accompanied by a covering letter from the Chairman of the KVI, Prof. Dr. Krupp, as well as a franked return envelope.

With few exceptions, the response to the questionnaires was very sluggish and only came after reminder letters were sent out. Answers from the unemployment benefit administrations are still lacking from France, Great Britain, and Austria. In a few cases (Austria and the Netherlands), the questionnaires were passed on from the administrative office that was contacted to the respective supervisory authority. Generally, the open questions were only occasionally answered, and then usually with considerable brevity. Relevant publications were rarely included with the responses.

5 National Models and Solutions: a Comparative Synopsis

In the countries selected here for examination, the research community and the official statistical offices have learned to work together in more or less systematic ways since the 1960s. In each case, institutionalized forms of cooperation have developed against a backdrop of country-specific legal and political frameworks, and cooperation has developed in conjunction with how research itself has been organized and institutionally channelled. As a memorandum on research in Switzerland and its relation to the Federal Statistical Office put it: “Intensified cooperation requires a minimum of institutionalization“ (Haug et al. 1998: 402). How statistical offices and the research community have organized the various tasks associated with data dissemination, and in particular which functions institutions within the research community exercise, is summarized below. The key aspects are access to data, data dissemination, costs and financing, and user consultation and support.

In terms of data access, or in terms of the official microdata made available, one should note that Scientific Use Files are standard products which need to be usable for a broad variety of research purposes. To ensure that this „collective good“ can be used in myriad ways, researchers both need to be in agreement and need to aggregate their various individual interests so as to be able to represent their interests collectively to the data producers. Some countries facilitated this process through institutions in the research community, especially where such institutions have particular data expertise and are in steady communication with their users through user groups or scientific advisory boards. Examples of this can be found particularly in Great Britain, the Netherlands, and in Norway. The discussion in Canada about the data to be made available through the Data Liberation Initiative takes place in the External Advisory Committee, half of whose members are from statistical offices and the other half of whom are researchers. Denmark and Sweden are the only countries in our survey where Scientific Use Files are not made available as a standard product of the statistical offices but are instead produced in accordance with the specific needs of a given research project.

The experience in the various countries has been that the statistical offices, as the data producers charged with the particular responsibility of ensuring data privacy as well as concerned with the willingness of those surveyed to provide information, tend to interpret confidentiality and anonymity guidelines narrowly. The possibilities laid out in law to also satisfy the data needs of researchers have often only been implemented once researchers made constructive suggestions about how files could be conceptualized or have suggested

procedures for making data anonymous. It is also true that making microdata anonymous always means reducing analytic possibilities, at least compared to working with the original data. The statistical offices do also have an interest in taking user interests into account and in making good data analysis possible with the anonymized data they supply. To this end a variety of informal means to exchange information between researchers and statistical offices can and do develop. But to ensure that the analysis interests of researchers are fully taken into account in the selection of what are legally mandated data protection measures, an institutionalized cooperation with the research community is more suitable than a mere cultivation of informal contacts, no matter how helpful such cultivation might be to both sides.¹⁰⁸

In light of the always meager means for financing statistical enumerations, it is also important to find thrifty ways to deal with the available resources. The goal of the broadest possible use of official surveys, where data is already collected using public monies, is served by secondary analysis conducted by researchers. Not only does this save on fiscal outlays that would be necessary in order to gather new data, it also contributes to the better analytic exploitation of the data collected. How valuable these data are, both for official statistical purposes and to researchers, can hardly be better demonstrated than through their varied and diverse use.

A pilot project in Germany that focused on certain specific microdata was able to increase the research use of this data by lowering the major barrier to access, namely the high costs researchers previously incurred in obtaining this data. The cost decrease was made possible because the Federal Ministry for Education and Research (BMBF) financed the preparation of the data, but this pilot project is slated to end in 2002/2003. Knowledge of financing models used in other countries, where access to official microdata is at issue, can provide answers to how the problems of high cost can be addressed in Germany once this pilot project comes to an end. This could contribute to the development of a new financing model to be applied to other (and future) official data.

¹⁰⁸ Germany has had a more or less formalized cooperation for a good while. A supervisory body, on which representatives of the federal data protection agency, the Institute for Employment Research [Institut für Arbeitsmarkt- und Berufsforschung; IAB] and researchers sat, provided guidance on data anonymization for the IAB-Employment Sample. A researcher-led body was established to provide technical advice to the Federal Statistical Office on anonymization, dataset creation, and the drafting of user friendly documentation on the Time Budget Survey, as part of the pilot project. The model for preparing properly anonymous microcensus data since 1989 was developed cooperatively between ZUMA and the Federal Statistical Office.

The selected countries differ not only in terms of the cost calculation principles each statistical office uses but also in how the service institutions in the academic research community are financed.

In most cases, the statistical offices calculate only the additional costs incurred in making the data available. In the USA, however, making microdata available is regarded as a public good and as part of the ordinary duty statistical agencies have to publish their findings. As a result, data are available at very moderate cost, and in part are even provided for free on the Internet. A similar understanding has developed in Great Britain and Norway, where official microdata are also passed on to researchers at virtually no cost. But in both Canada and the Netherlands there have been financing problems. In these two countries, starting in the 1980s, the statistical agencies could no longer count on public monies to completely cover the costs of collecting data itself. They were thus forced to calculate proportionate shares of data acquisition costs and levy them as part of the fees users had to pay to acquire the data.

As for service institutions within the research community, one can distinguish between the European and North American models. In Europe, with the exception of Denmark and Sweden, data service institutions and in some cases even data acquisition costs are typically financed through centralized or national research funds. No standard files are made available to researchers in Denmark and Sweden, and data fees are calculated by the statistical offices for each individual researcher or research project. In North America, the data infrastructure is largely financed through subscriber or member fees, in part reflecting different traditions of (decentralized) state administration and (partly privatized) research support (at least in the USA).¹⁰⁹

Given that the pilot project in Germany will soon come to an end, and given that there is a desire to ensure that official microdata will continue to be broadly accessible and usable, from the viewpoint of an individual researcher there are three viable models the research community could adopt to finance access to data:¹¹⁰

¹⁰⁹ The university-based financing model used in the USA is worth noting, since compared with European models far less public monies are used and financing is largely provided through grants and student fees.

¹¹⁰ In a survey of those who signed a memorandum on statistics initiated by GESIS, most researchers answered the question on the maximum amount they could spend on data fees by stating an amount ranging from 800 to 1,000 DM per file. The question referred to the means available to them as professors, without having to apply for external support. The pricing model of the federal and state statistical offices (in force since 1998), projects fees for anonymized microdata ranging from DM 200 to 4,000, depending upon the extent of the data. Thus, the estimated preparation cost for a microcensus, based on the mean number of requested variables, would be about DM 2,000.

(1) *Centralized financing through research funding agencies*, as it is practiced in Great Britain for Census data. Because the Economic and Social Research Council covers the costs, individual researchers pay little or no user fees.

(2) *Financing through subscription*, as it is practiced in Canada through the Data Liberation Initiative. Subscription fees paid by the universities cover 70 percent of the DLI costs. Researchers and students affiliated with the subscribing universities have free access to all Public Use Files from Statistics Canada.

(3) *Initial financing through research support and partial cost recovery through use-dependent fees*. The Dutch research community NWO [Netherlands Organisation for Scientific Research] uses this model to finance the estimated costs the statistical office incurs for the preparation of microdata and some of the estimated costs of the Scientific Statistical Agency (WSA). The file-specific data fees paid by the faculties or the research institutions are used largely to cover the WSA's costs.

In all three countries, these financing models have substantially contributed to a broad utilization of official microdata and have thus achieved their purpose. For an individual researcher searching for the least expensive and least bureaucratic access to data, the British model is the best. Not only is it characterized by low data costs but it also means there is open access to all interested university researchers. By comparison, data access in Canada is only possible if one is at a subscriber university, and access in the Netherlands requires a prior contractual agreement signed between a particular faculty and the statistical office.¹¹¹

Other models in which data is provided through service institutions within the research community should also be considered, as they may help to reduce the costs statistical offices incur in conjunction with disseminating microdata. Thus far in Germany, only the Scientific Use File of the IAB-Employment Sample from the Institute for Employment Research [Institut für Arbeitsmarkt- und Berufsforschung; IAB] within the Federal Labor Agency can be disseminated through a research community institution. Under Par. 16 of the Federal Statistics Act, anonymized microdata from state and federal statistical offices can only be passed on by the statistical offices themselves, and these offices are the only ones empowered to conclude use agreements or meet data protection obligations.

¹¹¹ It is worth noting that the particular financing models adopted in Canada and in the Netherlands were necessary because the proportionate costs for data acquisition were included in the data dissemination fees charged by the statistical offices; this situation that does not exist in Great Britain or other countries.

Official microdata are accessible exclusively through the statistical offices in Canada, Denmark, Sweden, and the Netherlands, but in all the other countries there are alternate ways to gain access to the data through institutions in the research community. The statistical offices in Great Britain and Norway even regard supplying data through data service institutions of their academic communities as advantageous not least for reasons of cost, and have even transferred substantial responsibilities over to them. The creation of a data agency which would act in the interests of all researchers is still under discussion in France. The Austrian data archive WISDOM has no particular data rights that would be comparable to the scope accorded the UK Data Archive, but it can pass on microdata to interested researchers after receiving authorization from Statistics Austria. And in the USA, though access to Public Use Files from various statistical agencies is inexpensive and simple, the ICPSR data archive has taken on the role of chief facilitator and intermediary for researchers and for ICPSR-affiliated institutes.

From the researcher point of view, there are various advantages to having data be passed on in completely documented formats through service institutions oriented to research needs as compared with receiving data from the official statistical bureaus. When data documentation is provided that meets international research standards, data users can save considerable data preparation time. In addition, by creating standard research classifications, the potential use of the data can be increased, something which is also true in the case of specially prepared files or data made comparable over time.

The need for consultation over research-specific issues applies to the entire research process, from the preparation of information needed about the data itself and access to it, to questions about the appropriate operationalization of a question or even to sophisticated methodological inquiries as to statistical modelling. In this context, the German pilot project provided an important impetus in developing and trying out different models of how researchers and statistical offices could work together, in particular when designated researchers became active consultants for questions about official microdata.

In one respect, data documentation prepared in a user-appropriate manner is particularly important because the official statistics will be used for secondary analysis, which is to say that the researcher will lack background information about the origin and preparation of the data itself. On the other hand, not all researcher inquiries are directly related to the data; they may instead be content or methodology questions. Service institutions in the research

community can best provide consultation services when they not only have data expertise but themselves also conduct their own, if limited, research. In that fashion they can gain a closer understanding of the issues that arise in data evaluation, an understanding that goes beyond documentation or advice about methods.

All statistical offices do provide some kind of user support. But in the KVI user survey, researchers generally judged the documentation provided by the statistics offices, as well as the support they received from the specialized bureaus in the statistics offices, to be rudimentary. The restrictive access through the statistical office, and in some cases the requirement that data be analyzed in statistical offices themselves, was connected to limitations on use in the view of the respondents. Consultative services and knowledge transmission, in the countries investigated, are far more readily and satisfactorily provided by the institutions that are themselves part of the research community infrastructure. The British and Norwegian Data Archives are particularly outstanding in this context, with researchers in these countries judging them (in the KVI user survey) to be good or very good in data dissemination, research-specific support, and data documentation. With their offerings of newsletters, mailing lists, informational meetings, and research seminars conducted in conjunction with the statistical offices, such research community institutions satisfy not only the needs of researchers for consultation and user support but also provide feedback to data producers from those actually using the data that was made available. In this manner, they make an invaluable contribution to the closer cooperation between the research community and the statistical offices.

References*

- Alba, R./Müller, W./Schimpl-Neimanns, B. (1994): Secondary Analysis of Official Microdata. S. 57-78 in: Borg, I./Mohler, P.Ph. (Hg.). Trends and Perspectives in Empirical Social Research. Berlin: de Gruyter.
- Albert, L./Austin, D./Boyko, E. (1996): The Canadian Experience. S. 9-15 in: SARs Newsletter No. 7, May. Manchester: Census Microdata Unit, Faculty of Economic and Social Studies, University of Manchester.
- Amann, A. (1992): Der Mikrozensus – Informationsquelle für Sozialberichterstattung und Sozialforschung. Österreichische Zeitschrift für Statistik und Informatik 22(4): 337-342.
- Association of Public Data Users: WWW-Site <http://www.apdu.org>
- Austin, E.W./Rockwell, R.C. (1997): The Python That Swallowed the Pig: ICPSR's 1990 Census Data Project. Michigan. ICPSR Bulletin (February).
- Bauer, P./Hackl, P. (1999): Das Bundesstatistikgesetz 2000 aus der Sicht des universitären Statistikers. Österreichische Zeitschrift für Statistik 28(3): 97-100.
- Bing, J. (1996): Dataprotection in Norway. URL:
http://www.jus.uio.no/iri/rettsinfo/lib/papers/dp_norway/dp_norway.html .
- Bradburn, N.M. (1999): The Future of Federal Statistics in the Information Age. Journal of Official Statistics 15(3): 353-366.
- Bureau of Labor Statistics: WWW-Site Current Population Survey
<http://www.bls.census.gov/cps/cpsmain.htm> ; National Longitudinal Survey
<http://stats.bls.gov/nlshome.htm> .
- Bustros, J./Berigan, J. (1999): Access to Statistics Canada Micordata files: the Canadian experience. S. 151-157 in: in Statistisches Bundesamt (Hg.): Kooperation zwischen Wissenschaft und amtlicher Statistik – Praxis und Perspektiven. Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden. Forum der Bundesstatistik; 34. Stuttgart: Metzler-Poeschel.
- Carnegie Mellon Census Research Data Center: WWW-Site
<http://www.heinz.cmu.edu/census/>
- Cecil, J.S. (1993): Confidentiality Legislation and the United States Federal Statistical System. Journal of Official Statistics 9(2): 519-535.
- Center for Electronic Records (National Archives and Records Administration): WWW-Site
<http://www.nara.gov/nara/electronic/>

* *If not stated explicitly, access to WWW pages as of August 2000.*

- Centre for Applied Social Surveys (CASS). WWW-Site
http://www.scpr.ac.uk/cass/docs/fr_casshome.htm
- Centre for Longitudinal Studies, Institute of Education, University of London: WWW-Site.
 ONS Longitudinal Survey: <http://www.cls.ioe.ac.uk/Ls/lshomepage.html> ; CLS
 Cohort Studies: <http://www.cls.ioe.ac.uk/Cohort/cstudie.html>
- Centre National de la Recherche (CNRS). WWW-Site
<http://www.cnrs.fr/cw/en/part/partener.html> .
- Church, J. (1999): Statistics: a matter of trust. S. 62-71 in: Statistisches Bundesamt (Hg.):
 Kooperation zwischen Wissenschaft und amtlicher Statistik – Praxis und
 Perspektiven. Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden.
 Forum der Bundesstatistik; 34. Stuttgart: Metzler-Poeschel.
- Citteur, C.A.W./Willenborg, L.C.R.J. (1993): Public Use Microdata Files: Current Practices
 at National Statistical Bureaus. Journal of Official Statistics 9(4): 783-794.
- Commission National de l'Information et des Libertés (CNIL). WWW-Site
<http://www.cnil.fr/> .
- Conférence des recteurs et des principaux des universités du Québec (CREPUQ): WWW-Site.
 SHERLOCK, Cooperative Data Access System for Quebec University Libraries.
<http://sherlock.crepunq.qc.ca/empdescr.html>.
- Conseil National de l'Information Statistique (CNIS). WWW-Site.
<http://www.cnis.fr/index.htm> ; Legal framework: http://www.cnis.fr/ind_english.htm .
- Dale, A./Elliot, M. (1998): A report on the disclosure risk of proposals for SARs from the
 2001 Census. The Cathie Marsh Centre: Working Paper No. 5, March 1998;
http://les.man.ac.uk/cssr/2001sars/issn_1367.htm .
- Danmarks Statistik (1991): IDA – an integrated data base for labour market research. Main
 report. Kopenhagen.
- Danmarks Statistik (1999): The Fertility Database. Guide to extracting data from the Fertility
 database. Kopenhagen.
- Datatilsynet / The Data Inspectorate: WWW-Site; Act No 48 of 9 June 1978 relating to
 Personal Data, Filing Systems, etc. [<http://www.datatilsynet.no/arkiv/engelsk/LOV-ENG.html>]
- David, M.H., (1998): Killing with kindness: The attack on public use data. Paper presented at
 the Joint Statistical Meetings, 11 August 1998.
- de Guchteneire, P./Mochmann, E. (Hg.) (1990): Data protection and data access. Amsterdam:
 North-Holland.
- de Guchteneire, P.F.A./Timmermans, J.G.M. (1990): Data Protection and Research in The
 Netherlands. S. 73-91 in: de Guchteneire, P./Mochmann, E. (Hg.) (1990): Data protec-
 tion and data access. Amsterdam: North-Holland.

- Department for Social Security (1998): Family Resources Survey. Great Britain 1996-1997. London: The Stationary Office.
- Department for Social Security (1999): The Family Resources Survey. London.
- de Vries, J. (1990): Country Report Canada. S. 9-20 in: de Guchteneire, P./Mochmann, E. (Hg.) (1990): Data protection and data access. Amsterdam: North-Holland.
- de Vries, W. (2000): PUMS, Privacy, Paranoia. Bericht für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI). New York: UN Dept. for Social and Economic Affairs, Statistics Division. (Ms.)
- de Vries, W./Nobel, J. (1999): Statistik, Geheimnisse und Empfindungen. Oder: Ein kurzer Überblick über die lange und komplexe Geschichte, wie in den Niederlanden statistische Mikrodaten schließlich offiziell für externe Forschungszwecke zur Verfügung gestellt wurden. S. 158-175 in Statistisches Bundesamt (Hg.): Kooperation zwischen Wissenschaft und amtlicher Statistik – Praxis und Perspektiven. Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden. Forum der Bundesstatistik; 34. Stuttgart: Metzler-Poeschel.
- de Wolf, V. (1995): Procedures for Researcher Access to Confidential Microdata at the Bureau of Labor Statistics. Washington: Bureau of Labor Office
[<http://stats.bls/ogov/orersch/st/st950040.html>].
- Dosselaar, P.G.W.M. (1993): Between Scylla and Charybdis: Controlling Disclosure Risks whilst Preserving Information. S. 445-450 in: Eurostat (Hg.). Internationales Seminar über die Statistische Geheimhaltung. Themenkreis 9, Reihe E. Luxemburg.
- Duncan, G.T./Jabine, T.B./de Wolf, V. (1993): Private Lives and Public Policies. Confidentiality and Accessibility of Government Statistics. Panel on Confidentiality and Data Access. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council and the Social Science Research Council. Washington D.C.: National Academy Press.
- Engelage, C. (1999): Bundesstatistikgesetz 2000 - einige Anmerkungen aus supranationaler Perspektive. Österreichische Zeitschrift für Statistik 28(3): 101-112.
- ESRC (1998) Reply to “Statistics: A Matter of Trust”, Ref. 022 in: Office for National Statistics (Hg.): Statistics: A Matter of Trust. Compendium of Replies to the Consultation Exercise 24 February – 31 May 1998. London: ONS.
- Europäische Kommission (1999): MISSOC. Gemeinschaftliches Informationssystem zur sozialen Sicherheit. Soziale Sicherheit in den Mitgliedsstaaten der Europäischen Union. Stand am 1. Januar 1998 und Entwicklung. Luxemburg: Amt für Amtliche Veröffentlichungen der Europäischen Gemeinschaften.
- Eurostat (1993): Organisation de la Statistique dans les Etats Membres de la Communauté Européenne. Volume II - Rapports analytiques. CSO-GSS, 245-266. Luxemburg: OOEPEC.

- Eurostat (1997): Proceedings of the Seminar on the Use of Administrative Sources for Statistical Purposes. Luxembourg, 15-16 January 1997. Luxembourg.
- Fellegi, I./Wolfson, M. (1999): Towards Systems of Social Statistics – Some Principles and Their Application in Statistics Canada. *Journal of Official Statistics* 15(3): 373-393.
- Fichter, D. (1998): The Evolution of Data Library Service in Canadian Academic Libraries. *DLI Update* 3 (Winter): 4-7.
- Final Report of the Joint Working Group of the Social Sciences and Humanities Research Council and Statistics Canada on the Advancement of Research Using Social Statistics, December 1998
(<http://www.sshrc.ca/english/policydocs/discussion/statscanreport.pdf> ; 05/13/99).
- Flaherty, D.H. (1979): Privacy and Government Data Banks. An International Perspective. London: Mansell.
- Ganzeboom, H.: WWW-Site. Tools for deriving status measures from ISKO-88. <http://www.fss.uu.nl/soc/hg/ismf/index.htm> ; 02.06.2000.
- Government Statistical Service (1984): The Government Code of Practice on the Handling of Data Obtained from Statistical Inquiries. Presented to Parliament by the Prime Minister by Command of Her Majesty. June 1984.
- Government Statistical Service (1999): Annual report, 1998-1999. London.
- Government Statistical Service: WWW-Site. <http://www.statistics.gov.uk/index.htm>
- Government Statistical Service (GSS): Framework for National Statistics. 1st edition. Operational from June 2000 (<http://www.statistics.gov.uk>).
- Granato, N./Schimpl-Neimanns, B./Schmidt, S., 2000: Zur Situation des Datenzugangs in ausgewählten Ländern. Eine Auswertung der KVI-Umfragen bei statistischen Ämtern und Nutzern aus der Forschung. Mannheim.
- Haug, W./Buhmann, B. (1998): Die Zusammenarbeit von amtlicher Statistik und empirischer Forschung bei der Datenerhebung und Datenanalyse. S. 281-304 in: Haug, W./Armingeon, K./Farago, P./Zürcher, M. (Hg.): Statistik im Dienste der Öffentlichkeit. Bern: Bundesamt für Statistik.
- Holvast, J. (1999): Statistical dissemination, confidentiality and disclosure. S. 191-207 in: Eurostat (Hg.): Statistical Data Confidentiality. Proceedings. March 1999. Luxembourg.
- Humphrey, C.K./University of Alberta (1996): Exploring Data Liberation. Putting DLI Data to Use. WWW-Site <http://datalib.library.ualberta.ca/accolleds/workshops/qa.html>.
- INSEE (1993): Le Système Statistique Public Français. 1ère Partie: Présentation générale. Edition 1993. Paris.

- INSEE (1993): Le Système Statistique Public Français. 2ème Partie: Les Services statistiques des Ministères. Edition 1993. Paris.
- Inter-university Consortium for Political and Social Research (ICPSR): WWW-Site <http://www.icpsr.umich.edu/index.html> .
- Institut de Recherches sur les Sociétés Contemporaines (IRESCO). WWW-Site <http://www.iresco.fr/menu/frame.htm> .
- Institute for Social Research: WWW-Site Health and Retirement Study (HRS) , Asset and Health Dynamics Among the Oldest Old (AHEAD) <http://hrs.isr.umich.edu:1041/reslis.htm> ; Overview of the Panel Study of Income Dynamics <http://www.isr.umich.edu/src/psid/overview.html> .
- Institut National de la Statistique et de Études Économiques (INSEE). WWW-Site http://www.insee.fr/fr/carte/accueil_carte.htm .
- Jabine, T. B. (1993): Procedures for Restricted Data Access. Journal of Official Statistics 9(2): 537-589.
- Joint Centre for Longitudinal Research (JCLR). WWW-Site <http://www.cls.ioe.ac.uk/Research/jclr.htm> .
- Keller, W.J./Kooiman, P./Barendswaard, M./van Rooij/Stol, H.R./Willenborg, L.C.R.J. (1993): Some Policy Issues at the Netherlands Central Bureau of Statistics. Journal of Official Statistics 9(1): 19-32.
- Keller, W.J./Willenborg, L.C.R.J. (1993): Microdata Release Policy at the Netherlands CBS. S. 439-444 in: Eurostat (Hrsg.). Internationales Seminar über die Statistische Geheimhaltung. Themenkreis 9, Reihe E. Luxemburg.
- Köhler, S./Schimpl-Neimanns, B./Schwarz, N. (2000): Pilotprojekt zur Erleichterung der Nutzungsmöglichkeiten von faktisch anonymisierten Mikrodaten. Wirtschaft und Statistik (1): 30-37.
- Kooiman, P./Nobel, J./Willenborg, L. (1999): Netherlands Official Statistics 14(Spring): 21-25.
- Kraus, F. (1995): Towards an Infrastructure of Comparative Research on Europe. EURODATA Research Archive Newsletter No.1 (Spring) [<http://www.mzes.uni-mannheim.de/eurodata/newsletter/no1/database.html>].
- Kraus, F. (1998): Towards a Data Infrastructure for Socio-economic Research on Europe: Improving Access to Official Microdata at the National and the European Level. EURODATA Research Archive Newsletter No. 7 (Spring) [<http://www.mzes.uni-mannheim.de/eurodata/newsletter/no7/no7.html>].
- Kutzenberger, E./Richter, J. (1999): Bundesstatistikgesetz 2000 - Vorgaben, Motive, Ziele. Österreichische Zeitschrift für Statistik 28(3): 131-150.

- Kvalheim, V. (1999): Managing Data Access and Confidentiality – A Model for Access to Government Statistics. S. 141-150 in: in Statistisches Bundesamt (Hg.): Kooperation zwischen Wissenschaft und amtlicher Statistik – Praxis und Perspektiven. Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden. Forum der Bundesstatistik; 34. Stuttgart: Metzler-Poeschel.
- Laboratoire d'Analyse Secondaire et de Méthodes appliquées à la Sociologie – Institut du Longitudinal (LASMAS-IdL). WWW-Site http://www.iresco/labos/lasmas/accueil_f.htm .
- Lang, G. (1995): Public Statistics and Statistical Confidentiality in France. S. 53-60 in: Eurostat (Ed.). International Seminar on Statistical Confidentiality. Proceedings, 28.-30.11.1994. Luxemburg.
- Lang, G. (o.J.): Vocabulaire du droit de la statistique et de la protection juridique des données individuelles. INSEE. Série des Documents de Travail. 'Coordination Statistique et Relations Internationales', N° D9502. Paris: INSEE.
- LASMAS (1988): L'Analyse Secondaire au LASMAS (11/1988). Paris: CNRS-IRESO.
- LASMAS (1993): L'Analyse Secondaire au LASMAS (6/1993). Paris: CNRS-IRESO.
- Lestina, G.J./LaPlant, W.P./Gillman, D.W./Appel, M.V./ U.S. Census Bureau (1996): Technical Development of the Proposed Statistical Metadata Standard. Washington: Bureau of the Census [<http://www.census.gov/srd/www/abstracts/gjlasa96.html>].
- Levine, F.J. (1996): Support for Sampling in Census 2000. American Sociological Association: Footnotes, Vol. 24, No. 8, p. 2.
- Lievesley, D. (1998): Institutional, legal and financial conditions for a successful co-operation between official statistics and scientific research. S. 305-314 in: Haug, W./Armingeon, K./Farago, P./Zürcher, M. (Hg.): Statistik im Dienste der Öffentlichkeit. Bern: Bundesamt für Statistik.
- Madrian, B.C./Lefgren, L.J. (1999): A Note on Longitudinally Matching Current Population Survey (CPS) Respondents. Technical Working Paper 247. Cambridge, MA: National Bureau of Economic Research.
- Manchester Information and Associated Services (MIMAS): WWW-Site; Government & Other Continuous Surveys <http://www.mimas.ac.uk/surveys/> .
- Marsh, A. (2000): The DSS/PSI Programme of Research Into Low-income Families (PRILIF). The Data Archive Bulletin No 73(January): 5-9.
- Marsh, C./Dale, A./Skinner, C. (1994): Safe Data versus Safe Settings: Access to Microdata from the British Census. International Statistical Review 62: 35-53.
- Marsh, C./Skinner, C./Arber, S./Penhale, B./Openshaw, S./Hobcraft, J./Lievesley, D./Walford, N. (1991): The Case for Samples of Anonymized Records from the 1991 Census. Journal of the Royal Statistical Society, Series A, 154: 305-340.

- Marsh, C./Teague, A. (1992): Samples of anonymised records from the 1991 Census. *Population Trends* 69 (Autumn): 17-26.
- Mayer, K.U./Huinink, J. (1994): Lebensverläufe und gesellschaftlicher Wandel: Von der Kohortenanalyse zur Lebensverlaufsanalyse. S. 92-112 in: Hauser, R./Hochmuth, U./Schwarze, J. (Hg.): *Mikroanalytische Grundlagen der Gesellschaftspolitik*. Band 1: *Ausgewählte Probleme und Lösungsansätze*. Berlin: Akademie Verlag.
- McGuckin, R.H. (1993): Analytic Use of Economic Microdata: A Model for Researcher Access with Confidentiality Protection. S. 83-97 in: Eurostat (Hg.). *Internationales Seminar über die Statistische Geheimhaltung*. Themenkreis 9, Reihe E. Luxemburg.
- Michigan Retirement Research Center: WWW-Site <http://www.isr.umich.edu/>
- Middleton, E. (1995): Internationally Comparable Measures of Occupational Status. *SARs Newsletter* (6): 1-2.
- Müller, W. (1999): Der Mikrozensus als Datenquelle sozialwissenschaftlicher Forschung. S. 7-27 in: Lüttinger, P. (Hg.): *Sozialstrukturanalysen mit dem Mikrozensus*. ZUMA Nachrichten Spezial; 6. Mannheim: ZUMA.
- Müller, W./Blien, U./Knoche, P./Wirth, H. u.a. (1991): Die faktische Anonymität von Mikrodaten. Band 19 der Schriftenreihe *Forum der Bundesstatistik*, herausgegeben vom Statistischen Bundesamt. Stuttgart: Metzler-Poeschel.
- Münz, R. (1999): Österreichs Statistik und die Chancen der Informationsgesellschaft. *Österreichische Zeitschrift für Statistik* 28(3): 151-153.
- Mugge, R.H. (1993): Informed Consent in U.S. Government Surveys. *Journal of Official Statistics* 9(2): 345-360.
- National Center for Education Statistics. WWW-Site. Available Surveys <http://nces.ed.gov/das/htm/surveys.htm>
- Nobel, J. (1995): Data Confidentiality and Data Access - Practical and Legal Issues in the Netherlands. S. 207-213 in: Eurostat (Hg.): *International Seminar on Statistical Confidentiality*. Proceedings. 28.-30. Nov. 1994. Luxemburg.
- Nobel, J. (1996): A law on official statistics. *Netherlands Official Statistics* (Summer): 48-53.
- Nobel, J. (1999): Statistical confidentiality – „So scrupulously honest and so tight-lipped“. S. 123-149 in: van Maarseveen, J.G.S.J./Gircour, M.B.G. (Hg.): *A century of statistics – counting, accounting and recounting in the Netherlands*. Amsterdam: Stichting beheer IISG.
- Norwegian Social Science Data Services (1996): *Current Activities and Projects*. NSD Newsletter (special edition). Bergen (<http://www.nsd.uib.no/english/nsd/newsletter/newsletter96.pdf>) .
- Norwegian Social Science Data Services (1999): *Providing Global Access to Distributed Data through Metadata Standardisation – The Parallel Stories of NESSTAR and the DDI*.

- Working Paper No. 10, presented to the UN/ECE Work Session on Statistical Metadata. Geneva, 22-24 September 1999. Genf: Statistical Commission and Economic Commission for Europe;
[\[http://www.unece.org/stats/documents/1999/09/metis/10.e.html\]](http://www.unece.org/stats/documents/1999/09/metis/10.e.html)
- Olson, J.A. (1999): Linkages With Data From Social Security Administrative Records in the Health and Retirement Study. Social Security Bulletin 62(2): 73-85
[\[http://www.ssa.gov/policy/pubs/SSB/v62n2y1999/notLinkagesWith.pdf \]](http://www.ssa.gov/policy/pubs/SSB/v62n2y1999/notLinkagesWith.pdf).
- Office for National Statistics (1996): Maintaining the Confidentiality of Data. London.
- Office for National Statistics (1996): Pricing and Dissemination Policy. London.
- Office for National Statistics (1996): Guide to Official Statistics. London.
- Office for National Statistics (o.J. [1996]): Government Statistical Service Code of Practice on the Handling of Data Obtained from Statistical Inquiries. London: ONS [originally published as White Paper (Cmnd. 9270; 1984)].
- Office for National Statistics (1998): Harmonised Concepts and Questions for Government Social Survey - Update December 1997. London: ONS
[\(<http://www.statistics.gov.uk/harmony/harmonfp.asp> \)](http://www.statistics.gov.uk/harmony/harmonfp.asp).
- Office for National Statistics (1999): Government Statistical Service. Annual Report and Accounts 1998-99. <http://www.ons.gov.uk/data/ons/annrep.pdf>
- o.V. (1999): Preserving and Enhancing Data on Education. ICPSR Bulletin (Fall).
- Scientific Statistical Agency / Netherlands Organisation for Scientific Research (1996): WSA Catalogue 1995-1996. Den Haag: WSA / NWO.
- Schimpl-Neimanns, B./Kraus, F. (1996): Verfahren der Weitergabe amtlicher Mikrodaten durch Statistische Ämter in ausgewählten Ländern (Kurzstudie). Mannheim (Ms.).
- Schimpl-Neimanns, B./Kraus, F., 2000: Weitergabe von Mikrodaten der amtlichen Statistik im internationalen Überblick. Bericht für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI). Mannheim.
- Schmähl, W./Fachinger, U. (1994): Prozeßproduzierte Daten als Grundlage für sozial- und verteilungspolitische Analysen - Erfahrungen mit Daten der Rentenversicherungsträger für Längsschnittanalysen. S. 179-200 in: Hauser, R./Ott, N./Wagner, G. (Hrsg.): Mikroanalytische Grundlagen der Gesellschaftspolitik. Band 2: Erhebungsverfahren, Analysemethoden und Mikrosimulation. Berlin: Akademie Verlag.
- Silberman, R. (1999): Les sciences sociales et leurs données. Rapport à l'attention de Monsieur le Ministre de l'Éducation Nationale, de la Recherche et de la Technologie. Paris: LASMAS-IdL. [see also <http://www.education.gouv.fr/rapport/silberman/>]
- SSHRC (Social Sciences and Humanities Research Council of Canada). WWW-Site: Expanding Research Capacity in the Humanities and Social Sciences. Discussion

Paper. May 25, 1998

(<http://www.sshrc.ca/english/policydocs/discussion/cfipaper.html>)

Statistical Policy Office/Office of Information and Regulatory Affairs/Office of Management and Budget (1999): Checklist on Disclosure Potential of Proposed Data Releases. Prepared by Interagency Confidentiality and Data Access Group. An Interest Group of the Federal Committee on Statistical Methodology. Washington; (http://www.fcsm.gov/docs/checklist_799.doc)

Statistics Canada (1997): Report on the Data Liberation Board of Management on the First Year of Operation of the Data Liberation Initiative. Ottawa (<http://www.statcan.ca/english/Dli/Document/findli.htm>).

Statistics Canada (1998): Quality Guidelines; <http://www.statcan.ca/english/IPS/Data/12-539-XIE.htm> .

Statistics Canada. WWW-Site: Statistics Act (1971) <http://www.statcan.ca/english/about/statact.html> .

Statistics Canada Research Data Centres (2000): Handbook for Researchers Under Agreement with Statistics Canada (Draft – October 2000). WWW-Site: <http://www.socsci.mcmaster.ca/rdc/rdcinvit.htm> ; 26.10.00).

Statistics Finland (o.J.): The Finnish Longitudinal Census Data File. Helsinki.

Statistik Österreich. WWW-Site: Bundesstatistikgesetz 2000 <http://www.oestat.gv.at/institution/bundesgesetz.pdf> .

Statistisches Bundesamt (Hg.) 1999: Kooperation zwischen Wissenschaft und amtlicher Statistik – Praxis und Perspektiven. Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden. Forum der Bundesstatistik; 34. Stuttgart: Metzler-Poeschel.

Statistisk sentralbyrå / Statistics Norway: WWW-Site http://www.ssb.no/vis/english/about_ssb/main.html ; various documents; among others The Statistics Act of 1989; http://www.ssb.no/english/about_ssb/statlaw/statlov_en.html)

Strike, G. (1998): Data Library Services at the University of Manitoba. DLI Update 3 (Winter): 3-4.

Subcommittee on Disclosure Limitation Methodology (1994): Statistical Policy Working Paper 22 – Report on Statistical Disclosure Limitation Methodology. Prepared by Subcommittee on Disclosure Limitation Methodology, Federal Committee on Statistical Methodology. Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget. Washington: Bureau of Transportation Statistics [<http://www.bts.gov/ntl/DOCS/wp22.html>]

Sylvester, Y. (1996): The ONS and The Data Archive. The Data Archive Bulletin 63 (September): 17-18.

- Tate, P. (1999): Longitudinal data from the Labour Force Survey. Labour Market Trends (July): 375-379.
- The California Census Research Data Center (CCRDC): WWW-Site <http://www.ccrdc.ucla.edu/index.html> .
- The Data Archive. WWW-Site <http://www.data-archive.ac.uk/> .
- The Cathie March Centre. WWW-Site <http://les.man.ac.uk/ccsr/Default.htm> .
- Unicon Research Corp. WWW-Site. <http://www.unicon.com> .
- United Nations Economic Commission for Europe (UN/ECE). Statistical Division's WWW site: Handbook of Official Statistics. <http://www.unece.org/stats/handbook/>
- University of Alberta: WWW-Site Putting DLI Data to Use. A Computing Exercise <http://datalib.library.ualberta.ca/accoleds/workshops/qace1.html> ; 02.06.2000
- U.S. Census Bureau: WWW-Sites. Survey of Income and Program Participation [<http://www.sipp.census.gov/sipp/>];
 The Census Bureau announces expanded Research Data Center Program [<http://www.census.gov/cecon/www.rdcannounce.html>];
 The Center for Economic Studies [<http://www.census.gov/cecon/www/ces.html>];
 Research Opportunities at the Census Bureau, The ASA/NSF/Census Bureau Research Fellow Program [<http://www.census.gov/srd/www/fellweb.html>] .
- Wallace, M./Landman, C./Sperling, J./Buczinski, C./U.S. Bureau of the Census (1999): Integrated Information Solutions – The Future of Census Bureau Data Access and Dissemination. Working Paper No. 7, presented to the UN/ECE Work Session on Statistical Metadata. Geneva, 22-24 September 1999. Genf: Statistical Commission and Economic Commission for Europe [<http://www.unece.org/stats/documents/1999/09/metis/7.e.pdf>]
- Watkins, W./Boyko, E. (1996): Data Liberation and Academic Freedom. Government Information in Canada 3(2, Fall), <http://www.usask.ca/library/gic/v3n2/watkins2/watkins2.html> .
- Wetenschappelijk Statistisch Agentschap (WSA; Scientific Statistical Agency). WWW-Site http://wsa.gamma.rug.nl/index_uk.htm .
- Wiggins, R.D./Beedham, H./Francis, B./Goldstein, H./Hannavy, M./Harman, J./Leyland, A./Musgrave, S./Rasbash, J./Smith, A.F. (2000): Teaching Resources And Materials for Social Scientists. <http://tramss.data-archive.ac.uk>
- Willenborg, L./de Waal, T. (1996): Statistical Disclosure Control in Practice. New York: Springer.
- Zayatz, L./Massell, P./Steel, P. (1999): Disclosure limitation practices and research at the U.S. Census Bureau. Netherlands Official Statistics 14(Spring): 26-29.

**Appendix: Documentation of the survey of statistical offices and
service institutions of the research community**

1A) France:	
	Institut National de la Statistique et des Études Économiques (INSEE; Statistics France), Paris 69
1B) France:	
	Laboratoire d'Analyse Secondaire et des Méthodes Appliquées à la Sociologie - Institut du Longitudinal (LASMAS-IdL), Paris 73
2A) United Kingdom:	
	The Office for National Statistics (ONS), London 79
2B) United Kingdom:	
	Manchester Information and Associated Services (MIMAS), Manchester 83
3A) Canada:	
	Statistics Canada, Ottawa (DLI Unit) 87
3B) Canada:	
	University of Alberta, Data Library (Local DLI Contact) 90
4A) The Netherlands:	
	Centraal Bureau voor de Statistiek (CBS; Statistics Netherlands), Voorburg 96
4B) The Netherlands:	
	Wetenschappelijk Statistisch Agentschap (WSA; Scientific Statistical Agency), Den Haag 101
5A) Norway:	
	Statistisk Sentralbyrå (SSB; Statistics Norway), Oslo 108
5B) Norway:	
	Norsk samfunnsvitenskapelig datatjeneste (NSD; Norwegian Social Science Data Services), Bergen 111
6A) Austria:	
	Statistik Österreich (Statistics Austria), Wien 116
6B) Austria:	
	Wiener Institut für Sozialwissenschaftliche Dokumentation und Methodik (WISDOM; Vienna Institut for Social Science Documentation and Methods) 119
7A) U.S.A.:	
	U.S. Bureau of the Census and other federal statistical agencies 124
7B) U.S.A.:	
	Inter-university Consortium for Political and Social Research (ICPSR), Institute for Social Research, University of Michigan, Ann Arbor 130

1A) France: Institut National de la Statistique et des Études Économiques (INSEE; Statistics France), Paris

Michel Jacod*

1.1 Legal regulations concerning collection of data**

What is the legal basis for the compilation of official microdata in your country: general framework legislation (e.g. statistics law) and/or single survey legislation?

General framework legislation (1951 act)

{1.2.3 Institution responsible for determining the program for data collection: CNIS Conseil National de l'Information Statistique: this national conseil for statistical information is solicited to appreciate the opportunity of each national operation. The final programme is determined by INSEE and various statistical offices within the various administrations, based on the CNIS' conclusions.}

1.2 Legal regulations concerning the dissemination of microdata

{3.4 Eligible for access to scientific use files:

Universities and public non university research institutes.

3.5 Restrictions on use of files:

Specified research project.

Applies only to public microdata files of the social field: 4.1.1.6 Contract for data use / 4.1.1.7

Period of use / 4.1.1.8 Restrictions on specific purposes:

No Contract. Opening the file means acceptance of an end user license without any allowance for re-dissemination of any kind. / Unlimited period of time / No restrictions.}

2.1 Released microdata files

Do you disseminate microdata also directly to individual bona fide researchers?

Yes

If yes: Which surveys on households and individuals are disseminated?

Essentially the labour force survey, but every survey file can be accessed on request and after anonymization.

* Michel Jacod is Director of the Department for Data Dissemination and Regional Affairs of the INSEE.

** Information from an earlier report (Schimpl-Neimanns/Kraus 1996) was supplemented by replies of the INSEE to a KVI survey, made available to us by the Federal Statistical Office of Germany. These replies are indicated by curved brackets {}.

2.2 Formats of released microdata

In which form are data distributed (scientific use files, user-tailored files) and which are the conditions for usage?

Public Use Files. No specific conditions unless on data property (no redissemination allowed).

In which formats (ASCII/EBCDIC, SAS, SPSS) are data distributed?

ASCII, SAS, DBF, Beyond 20/20

2.3 Data documentation

What type of documentation is provided (codebook, questionnaire, other)?

As far as possible, integrated documentation (contextual pop in) documentation integrated to the Beyond 20/20 files, facsimile of the questionnaire and more eventually.

2.4 Consultative services

Does INSEE provide consultative services to academic data users? If yes: In which form and to which extent? Advice on use and meaning of data.

Yes, as to the other users.

2.5 Organization of the data distribution within the statistical office

Is there a central unit established within INSEE for distributing data to academic research?

No, the general dissemination network of INSEE (28 sites).

2.6 Scope of methods for disclosure control

To which extent are measures taken to anonymize data (combining of categories, removal of variables, identifiers, etc.)

Removal of all identifiers (especially geographic codes under the „département“ level)

{3.1.1: Microdata as public use files:

Sample survey: Removing of direct identifiers and geographic codes under the “department” (100 departments in France).

Census data on individuals: no geographic codes under 50,000 inhabitants.

Census data on housing: no geographic codes under 2,000 inhabitants (fixed zones of about 2,000 inhabitants).}

3. Release of microdata to data archives and/or individual researchers [see also item no. 2]

To which academic organisations – other than LASMAS - are microdata disseminated for further distribution to affiliated researchers and under which conditions?

No one

4. Rights of data users

{ 1.3.2 Requirements concerning access to the microdata:

No precise and official regulation can be edicted on a general basis. A distinction should be made between business data and personal or household data.

In this second area, rules are established source after source, after consulting the CNIL (independent administrative body in charge of implementing the 1978 privacy law), these rules frequently distinguish governmental users who are allowed access to more detailed data from the users; in this regard, researchers cannot be qualified on an a priori basis, and the access to microdata which are reserved to governmental users, constitutes a specific case to be instructed by the CNIL.

Business microdata provision to researchers is regulated by a subcommittee of the CNIS “secret committee” where researchers are represented, but several other interest groups, especially business organisations in order to balance scientific interests with business confidentiality. }

5. Cost

On which basis are total costs calculated (partial refunding of costs for data collection and data anonymization, charge for documentation costs, charge for dissemination costs)

Recovering costs for dissemination including documentation, anonymization, formatting, marketing.

What is the price range charged to individual researchers (cheapest price/name of survey and highest price/name of survey)?

3 ranges of price depending on the number of variables and the frequency of the survey, between about 1000 Euro and 2000 Euro.

{ 4.1.1.9 Costs for the user:

Three price levels, depending on the number of variables in the survey and on its periodicity because prices reflect the dissemination costs (and not the production costs) which are strongly related to the specific documentation work: more variables = more documentation to be written, more frequent survey = more previous documentation re-usable.

Other costs to be recovered = anonymization, support, infrastructure (including marketing costs).}

6. Experiences of the statistical office

Which are the experiences made by INSEE regarding the co-operation between academic research (CNRS research units/individual researchers) and your organisation? (relevance of academic research, adding of value through research, violation of usage conditions, etc.)

Very good co-operation in case of partnership for conception or primary analysis related to specific surveys. Not very efficient in other cases: difficulty for getting organised feedback from researchers.

{4.1.1.13 Demand for public and scientific use files:

About 20 files sold during the 2nd half of 1999 (the pricing policy was drastically changed in June).

4.1.2: Grants:

Discussion are about to be finalised with the French research administration. A general agreement could be passed in the coming months to open two research data access centers able to collect microdata files, to document them if useful (that would be the case for files produced by some ministry statistical offices which have difficulties in this field), to provide scientific users with files, documentation and advices. A scientific committee would select the valuable official statistical files and the legitimated research projects. A financial contribution would be paid by scientific users, but it would be less expensive than that required in case of direct buy to INSEE.}

1B) France: Laboratoire d'Analyse Secondaire et des Méthodes Appliquées à la Sociologie - Institut du Longitudinal (LASMAS-IdL), Paris

Roxane Silberman und Irène Fournier Mearelli*

1. Conditions of data acquisition**

Which possibilities does the scientific community generally have in your country to use official microdata, apart from requesting data from your organisation? If access limited to certain types of data only: survey data on persons and households; survey data on enterprises: register data (e.g., social security records, etc.)?

LASMAS does not monopolise the access to archived datasets. In most cases, and particularly for INSEE datasets, researchers can access the datasets directly. On the other hand, there are some advantages to gaining access through LASMAS:

- User fees for the datasets are paid at a flat rate for all researchers. By contrast, researchers or research institutes that obtain datasets from INSEE directly must bear higher costs.
- The data is provided more quickly.
- Documentation is provided in standardised form.
- Support and assistance as to the structure of datasets, about variables, or for other questions are more readily available through LASMAS.

The three last points are also the reason why INSEE itself wanted LASMAS to handle data access. Also, there is no longer privileged access to the datasets held at the BDSP [French data archive for academic survey data] However, the datasets archived at LASMAS and at the BSDP do not cover the entire universe of microdata by far.

In this context there are no clear rules. Data producer practices regarding publicly accessible data varies and is strongly dependant upon the actual relationship a researcher has to the relevant statistical department. Access is often limited to the research work carried out that has been ordered by the client. Recent reports, produced in conjunction with the utilisation of publicly accessible data, have more explicitly addressed making data available to a broad public than they have discussed researcher access to microdata. In such reports, suggestions have been made that at least partly mirror LASMAS or BDSP practices.

In particular cases, access to datasets is regulated differently:

- The approval of the CNIS Data Protection Committee is necessary for access to the Survey of Firms.
- For access to administrative data related to individuals, the respective administrative unit can provide the necessary approval. Access is overseen by this administrative unit, and they must in turn be able to justify it to the CNIL oversight body.

* Roxane Silberman is Director of LASMAS-IdL, Irène Fournier Mearelli is staff member of LASMAS-IdL.

LASMAS, however, is in negotiation over an agreement which would give it responsibility for access to DARES (an Employment Ministry office) datasets. In a similar manner, access to data generated in the administrative court system is being organised through research programs; the data would come from GIP (in the Justice Ministry).

For access to Census data at the sub-community level (the threshold level was recently increased by the CNIL oversight authority), a protected area is to be created that will permit researchers access (though under CNIL supervision). There are plans to have LASMAS be involved in conceptualising the organisation of this protected area.

According to our information, LASMAS serves official microdata to members of CNRS research units (permanent or associated). What does this mean in terms of coverage? (What are the conditions for permanent resp. associated membership and what is the approximate coverage of French social scientist?)

LASMAS at present organises access for all researchers and teachers who work at (or are assigned to) a CNRS-associated research facility, inasmuch as they are working with INSEE data.

In other cases, access is more broadly regulated; LASMAS is empowered to make datasets accessible to the entire research community and the disseminate them (CEREQ, Ministry of Culture)

In the former case, based on estimates, about half the researchers are not associated with the CNRS. There are a growing number of specialised institutes (such as INED, INRA, or INREST) which have concluded more or less advantageous user agreements with INSEE, as well as graduate programs at universities which have numerous doctoral students enrolled, but for both groups - which are growing - access to data through LASMAS is impossible. The costs of access to the data for these research institutions are prohibitive.

The possibility that LASMAS will make data accessible to all researchers is now under discussion.

Under which conditions can your organisation obtain official microdata? Which are the basic features of, e.g., the contracts concerning the transfer of data to your organisation?

The use of data for research purposes has been established as a basic principle. The goal is to expand the utilisation of the datasets that have been created, itself connected to a corresponding increase in resource needs, and thereby ease the financial burden on the data producers who are currently providing costly user services.

In return, LASMAS obligates itself to make the datasets accessible under the best possible conditions, which includes documentation preparation and user support. LASMAS also will be

responsible for providing feedback about data use (completed studies) back to the data producers, and thereby contribute to improved data quality.

For data on enterprises archived at DARES, there is a special use proviso which is based on upholding professional moral and ethical standards as well as principles of confidentiality. LASMAS ensures that researchers understand and sign a declaration of obligation in which they accept the principles of access and use.

The current goal is to widen access to additional data sources by means of professional self-regulation (including through a scientific oversight committee), as well as to improve the feedback provided to the data producers.

2. Acquisition concept of the data archive

Which official microdata are requested by your organisation? Which criteria is the selection of data, if necessary, based on?

Initially, the selection criterion was bound to the research area assigned to LASMAS. Later, user requests and wishes from researchers at other research centres were also taken into consideration. Then the effort was made to complete the longitudinal datasets. The objective is to create a scientific committee that will be given responsibility for decisions about data acquisitions that will be based in part on user recommendations (for a current list of the available data, see <http://www.iresco.fr/labos/lasmas/enquetes.htm>).

3. Data formats, documentation, consultative services

In which form (raw data, portable system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata transferred?

Normally in SAS format. However, the datasets are in a “transit format” that makes their conversion possible, as they are SPSS-readable. They will soon also be available in ASCII format.

We are also trying, little by little, to provide the documentation online. For this we need an overview of the surveys, lists of variables, variable codes, some of the counts, and the bibliography. Most of the documentation still exists on paper.

How are the data processed by your organisation, is additional documentation provided?

For the moment, not yet.

Which consultative services does your organisation provide to academic data users with respect to official microdata, to which extent and in which form?

Variable extraction, preparation of written user aids (though this needs more development).

4. Rights concerning data usage

Can official microdata be distributed to third parties, researchers, institutes, etc.? Under which conditions? Are there special conditions regarding access to longitudinal data and census data?

We do not have panel data. It is, however, possible to create artificial longitudinal datasets by combining cross-sectional data.

There are no particular conditions, or exceptions to the regulations, that INSEE imposes for the data provided, up to and including the 1990 Census. As for the future, we are currently negotiating with INSEE. The terms and conditions of access will be somewhat more restricted, particularly for regional data from areas with less than 50,000 inhabitants, and for sensitive data about foreigners.

5. Modes of data access

What do researchers who are interested in official microdata have to do to obtain copies of data or to use them? What is the role of your organisation in this case?

See the declaration of obligation available on the World Wide Web. The main purpose for the use of data must be research. We make excerpts available (select variables, individual population groups) as well as the entire dataset.

In the case of doctoral students it is necessary that the supervisor in charge also sign the declaration of obligation.

There are no difficulties in the case of longitudinal data. The data are made available to us in anonymized form upon request to INSEE, though there is no community-level information provided.

Which are the major rights and obligations of data users?

Research must be financed out of public monies, and the results must have no limits placed on them that would violate the right to publish. Users personally obligate themselves, in writing, to not give the datasets to third parties, and also to protect access to or use of the data by those who do not have access or data use permission. They also obligate themselves to identify the data source when the results are based on an analysis of the data. Finally, users promise to send LASMAS-IdL a copy of every publication which used survey data that were made available by LASMAS-IdL. In addition, a brief description of how the data were used, along with what the user experiences were in working with the data, is to be handed in.

6. Acquisition cost

Which costs does the data provider (statistical office, register authority) charge to your organisation a) for the purpose of partial refunding of data collection costs, b) for the processing, etc., of data, c) for user services?

INSEE is financed through the national budget. It does not have the goal of passing the costs of data acquisition and dataset preparation on to its users.

For those surveys made available from INSEE, LASMAS (acting on behalf of CNRS) pays a flat rate data use fee for all researchers, thereby considerably reducing the costs to each user. Previously, those in the research community paid half-price compared to other users (such as private firms or local and regional organisations), but this reduced cost will be abolished as INSEE has substantially reduced microdata preparation costs (e.g., the cost of the employment sample has dropped from FF 40,000 to FF 8,000)

7. Fees charged by the data archive

How high are the costs charged by your organisation for data transfer, what are these costs composed of (costs for data, direct costs for the distribution of data, e.g. for data carriers, costs for the production of photocopies, for staff, a proportion of the costs/flat rate for the generating of data, data documentation costs, etc.)?

No fees are charged.

8. Financing

How is your organisation financed? Especially: who bears the costs that emerge for the acquisition of microdata if your funds do not cover the acquisition/distribution of that data?

CNRS bears the costs (or more precisely the SHS division); every year governmental budgetary allocations are made to cover the costs.

9. Demand for official microdata

How strong is the demand for official microdata, how extensive is the provision of consultative services?

About 200 analyses using the various surveys were conducted in 1999.

If possible, we include (at LASMAS's cost) a copy of the questionnaire, the coding scheme, and the instructions given to interviewers.

How is steering of your data services organised? (External boards, relation to IRESCO and CNRS)

We are a division in a CNRS research center, but are at two locations. In Paris, we are located at IRECSO, one of the institutes associated with the CNRS. In Caen we are at the university. The services offered are provided by 7 part-time employees; Roxane Silberman is the director. The research center is under the scientific direction of the "Social and Human Sciences" (SHS)

division in CNRS.

10. Experiences of the scientific community

Which are the experiences made by academic research concerning the co-operation with the Statistical Office? Where could improvements be made from the scientific community's point of view?

Both INSEE and the individual Ministries conduct their own research, using data they themselves have collected or have contracted out. There is no sharp division in France between data production and the conduct of research. There is also a long-standing close connection between the research community and the producers of statistics. This co-operation extends not only to the selection and conceptualisation of topics to be covered in surveys but also to data evaluation and publication.

Until a decade ago, individual researchers had access to INSEE data; this form of access is now becoming rarer. In part this is due to a more restrictive access policy for microdata, a consequence of a public today more sensitised to data privacy issues as well as owing to pressure from the CNIL, and in part it is due to increasing marketing the data by INSEE itself. As this change was occurring, CES and its successor LASMAS were creating up an archive for various INSEE-conducted surveys. This archive was financed as well as contracted by CNRS, and has successively acquired surveys on occupational qualification (Enquête FQP; since 1970), on leisure activity (Enquête Loisirs), and since 1986, the employment sample survey. Since then, the archival holdings of surveys has continued to increase. The institutionalised access through LASMAS, and the priority researchers have to the data, is regarded as advantageous by INSEE.

Our research center in this context is slated to play an increasingly important role in the future.

2A) United Kingdom: The Office for National Statistics (ONS), London *

1.1 Legal regulations concerning collection of data

According to which rules/norms are official microdata compiled, and is there a legal basis for it (statistics law, etc.)?

There is no overarching or framework law in the United Kingdom. The legal framework for the surveys instead is based on the passage of successive, individual laws. None of these laws distinguishes between enterprises and individuals: thus in principle both are subject to the same rules.

1.2 Legal regulations concerning the dissemination of microdata

According to which rules are microdata distributed to the scientific community?

Measures to ensure the confidentiality of statistical information as well as the rules for dissemination are laid out in a series of institutionalized Codes of Practice as well as in the Data Protection Act of 2000.

The Statistical Office is explicitly enjoined to stimulate research, to make data ready and available, and to provide consultative services. With very few exception, microdata are only disseminated in anonymous form and only once the Data Custodian has given his approval. The criteria for this are set forth in the Guidelines for Data Custodians.

One should distinguish here between two fundamentally different cases:

- Data whose confidentiality is potentially more at risk can, in the context of the legal regulations, be made available to bona fide researchers, conditionally, under strict controls, and limited in terms of contents and in the amount of time they can be used.
- Data that is anonymized, where the potential risk of reidentification is relatively low, can, in the context of the use by researchers, be passed on to the Data Archive of the British research community ESRC. The Data Archive then acts as a trustee for the ONS.

{ 1.3.2 Requirements concerning access to microdata:

The legal framework does not prevent the giving of access to anonymized microdata in which individual units cannot be identified.

3.4 Eligible for access to the scientific use files / 3.5 Restrictions:

Non-commercial users – mainly academics, research institutes, government departments. / Specified research projects }

3.6 Special features regarding access to scientific use files:

* Information from an earlier report (Schimpl-Neimanns/Kraus 1996) was supplemented by replies of the ONS to a KVI survey, made available to us by the Federal Statistical Office of Germany. These replies are

Those accessing the files must sign an agreement specifying the conditions under which access is given.}

2.1 Released microdata files

Which official microdata are distributed to the scientific community (e.g. Labour Force Survey, household budget, etc.)?

The most important surveys - Labour Force Survey, General Household Survey, Family Expenditure Survey, British Household Panel - are available through the Data Archive at the University of Essex. Samples from the last Censuses have been available online since the early 1990s through the University of Manchester's MIMAS (Manchester Census Micro Data Unit), an entity organised especially for this purpose.

{3. Microdata offered by OSN:

No.

{4.1. Does ONS grant access to the microdata itself.

Not granted by ONS.

4.1.2.1 Organization responsible for access to the microdata of ONS:

The Data Archive, a unit funded by the Economic and Social Research Council, the central government body for funding research. The unit is based at the University of Essex.

4.1.2.12 ... Advantages for the users being granted access by an external institution?

The Data Archive is a specialised institution which holds microdata from a large number of sources. It has economies of scale.}

2.2 Formats of released microdata

In which form (raw data, system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata distributed?

Data from the Statistical Office are transformed by the Data Archive so as to be usable by the most common statistical analysis packages (SPSS, SAS, etc.), and made available to interested researchers (for Census samples, see MIMAS, Appendix 2B).

{4.1.2.2 – 4.1.2.9 Format of the data – costs for the users:

(...) These are mainly matters which are decided by The Data Archive.}

2.3 Data documentation

The Data Archive prepares data documentation of the Statistical Office according to research-specific criteria.

indicated by curved brackets {}.

2.4 Consultative services

Does the Statistical Office provide consultative services to academic data users; if it does: to which extent and in which form?

Consultation services are provided particularly through the support given to user education (through user groups of the Data Archive and meetings of the Royal Statistical Society).

2.5 Organization of the data distribution within the statistical office

Is there a central unit within the Statistical Office responsible for distributing data to academic research?

The Statistical Office does not provide services to individual researchers, and instead deposits the data at the Data Archive. Inquirers are always directed to contact the Data Archive.

2.6 Scope of statistical methods for disclosure control

To which extent are measures taken to anonymize data (combining of categories, removal of variables, identifiers, etc.)?

Up to this point, four major procedures have been followed: the drawing of subsamples, the suppression of variables, the grouping of regional identifiers, and the grouping of response categories.

{3.3.1/3.3.5 Anonymization measures (survey data on households and persons / other survey data):

Removing direct identifiers, recoding.

3. Release of microdata to data archives and/or individual researchers

Are the data distributed to a data archive or other central scientific institution for further transfer to researchers/research institutes or only to individual researchers/research institutes upon request?

Microdata (in the form of Scientific Use Files) released is, as a matter of principle, deposited by the ONS for research use at the Data Archive. The Data Archive passes this data on in part to specific institutions within the British research community. The Census sample is available through the Census Micro Data Unit (MIMAS) (see Appendix 2B). The Data Archive functions in a trusteeship capacity relative to users in the research community.

A contract is concluded with the individual or institution for all data requested, with the Archive acting as a trustee; requests made to the Data Archive must include a research plan.

4. Rights of data users

Which rights does the data recipient acquire? Can a data recipient, for example a data archive, transfer data to third parties; if this is possible: which are the conditions; or does the data recipient have to use the data exclusively for himself/herself?

Access to other microdata from official statistics comes about with the concluding of a contract with the Data Archive. Use is restricted: those acquiring the data can only permit access to this data by third parties with the explicit permission of the Data Archive. This is true not only for the microdata itself but also for the tables or statistics that are created or calculated based on these microdata files. The contract does not limit how long the data can be utilised, nor is there an obligation to destroy the data.

5. Cost

Which are the total costs charged by the Statistical Office for the distribution of data, what are these total costs composed of (costs for data, direct costs for data distribution, e.g. for data carriers, costs for the production of photocopies, for staff, a proportion of the costs of/flat rate for the generating of data, data documentation costs, etc.)? If data are supplied to a scientific institution, for example a data archive, which costs are charged to the final user?

{4.1.2.10 ... who bears the remaining costs of offering the data (for example, for anonymizing the data or for infrastructure)?

The Data Archive is funded by the Economic and Social Research Council to encourage and support social research by allowing the maximum use to be made of data collected in governmental surveys. The cost of anonymizing microdata is borne by ONS. }

6. Experiences of the statistical office

Which are the experiences made by the Statistical Office regarding the transfer of data to academic research? (With respect to the usability of research results, how discreet data are handled by scientists, research co-operation between the Statistical Office and academic research, etc.).

See, among others, Sylvester (1996); Church (1999).

**2B) United Kingdom: Manchester Information and Associated Services (MIMAS),
Manchester**

Keith Cole *

1. Conditions of data acquisition

Which possibilities does the scientific community generally have in your country to use official microdata, apart from requesting data from your organisation? If access limited to certain types of data only: survey data on persons and households; survey data on enterprises: register data (e.g., social security records, etc.)?

The UK academic community has excellent access to the Census of Population statistics and the major government surveys and other continuous surveys.

The Census and Survey datasets held at MIMAS fall into two main categories: Aggregate and Individual level. The aggregate datasets are predefined cross-tabulations of one or more variables. For example, the Census Area Statistics (CAS) are a set of pre-defined tables for small geographical areas (approx. 100 households). These tables are subject to various confidentiality techniques to prevent inadvertent disclosure of information about individuals. The individual level datasets provide access to data for individual households and/or individuals. Various techniques are used to ensure that it is not possible to identify individuals and/or households.

Which is the major division of labour between The Data Archive and MIMAS?

MIMAS is a JISC (<http://www.jisc.ac.uk>) [Joint Information Systems Committee] funded national data centre and provides flexible on-line access to a range of bibliographic and non-bibliographic data and information resources. See <http://www.mimas.ac.uk>. Part of the MIMAS socio-economic data service (i.e. access to the large government and other continuous surveys) is run in conjunction with the Data Archive. MIMAS is responsible for providing secure on-line access, facilitating access and providing content 3) level support.

Under which conditions can your organisation obtain official microdata from the Statistical Office? Which are the basic features of, e.g., the contracts concerning the transfer of data to your organisation?

Most of the data held by MIMAS is negotiated by other agencies - including the Data Archive.

* Keith Cole is MIMAS Service Manager and head of the Census Dissemination Unit. Reply on behalf of the UK Data Archive to which the questionnaire originally was addressed to.

2. Acquisition concept of the data archive

Which official microdata are actually requested by your organisation?

In general terms, MIMAS does not have a role in deciding what microdata should be obtained for the academic community. All data requests are passed on the appropriate body - such as the Data Archive, ESRC or JISC.

Which criteria is the selection of data, if necessary, based on?

In general terms, all data acquisitions need to be justified on the basis of potential use in teaching and research. MIMAS only holds datasets for which there is a large potential demand.

3. Data formats, documentation, consultative services

In which form (raw data, portable system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata transferred to your organisation?

A number of the government and other continuous surveys come as SIR export format files which MIMAS loads as SIR data bases. Some of the surveys, such as the BHPS [British Household Panel Study] come in a variety of formats (SPSS, SAS, STATA & SIR). These formats usually have embedded metadata in the form of variable names, value labels etc. Many of the Census datasets come as formatted ASCII files with little embedded metadata.

How are the data processed by your organization, is additional documentation provided?

Data processing tasks principally include loading the data into an appropriate software system (e.g. SIR etc). Some Quality Assurance is often required due to format changes etc. MIMAS endeavours to provide value added services, such as documentation, training and user support. This includes making information about how to access the data and other information available through the web (see <http://www.mimas.ac.uk/surveys> and <http://census.ac.uk/cdu>). Some interface (e.g. web) development work is also done at MIMAS.

Which consultative services does your organisation provide to academic data users with respect to official microdata, to which extent and in which form?

MIMAS provides a range of support services. Two specialist support staff provide content level support on the Census and Surveys dataset services.

4. Rights concerning data usage

In which ways can the official microdata transferred to your organisation be accessed by third parties and which are the major conditions for access?

In general terms we are not allowed to pass on data to any unlicensed third party. Providing a third party is licensed it is possible for data to be transferred. However, MIMAS is not funded to provide services to non-academic bodies.

5. Modes of data access

What do researchers interested in official microdata have to do to obtain copies of data or to use them via other modes of access? What is the role of your organisation in these cases, and how do you ensure feedback between data users and data collectors?

All users of the Census and Survey datasets held at MIMAS must be registered and have agreed to the terms and conditions of use. For the survey datasets, users register with the Data Archive. For the Census data sets they register with the appropriate registration agency - of which there are a number. MIMAS will grant access as soon as it has received authorisation from the appropriate registration agency.

Which are the major rights and obligations of the data users?

Essentially, only to use the data for teaching and academic research; not to pass the data onto any unlicensed third party and not to use the data to identify individuals. Some agreements also require users to report back on use.

6. Acquisition cost

Which costs does the data provider (statistical office, register authority) charge to your organisation? For the purpose of partial refunding of data collection costs / for the processing, etc., of data (anonymization etc.) / for user services?

N/A see (10)

7. Fees charged by the data archive

How high are the costs charged by your organisation to a data user and what are these costs composed of (refund of data acquisition costs; refund of your organisation's costs for data processing; costs for the distribution of data, e.g. for data carriers; data documentation costs, refund of staff costs)?

There are no charges to end-users of the MIMAS socio-economic data service. The costs of running the service at MIMAS are met by JISC and ESRC.

8. Financing

How is your organisation financed? Especially: who bears the costs that arise for the acquisition of microdata if the data archive's funds do not cover the acquisition/distribution costs of these data?

MIMAS is funded by the Higher Education Funding Councils through JISC and also receives some funding from ESRC for the Census service. The costs of the datasets held at MIMAS are usually borne by other agencies (e.g. JISC and ESRC).

9. Demand for official microdata

How strong is the demand for official microdata, how extensive is the provision of consultative services?

In general terms there is a growing demand within UK Higher Education for all types of data and information resources. However, existing microdata resources are probably not fully exploited in teaching and research as much as they could be due the lack of appropriate data analysis skills

10. Experiences of the scientific community

Which are the experiences made by academic research concerning the co-operation with the Statistical Office? Where could improvements be made from the scientific community's point of view?

There is considerable consultation between the Census Offices and the academic community over the forthcoming 2001 Census and this is valued by both sides. The problem with all consultation is engaging with the wider user community.

3A) Canada: Statistics Canada, Ottawa (DLI Unit)

Jeanine Bustros *

1.1 Legal regulations concerning collection of data

According to which rules/norms are official microdata compiled, and is there a legal basis for it (statistics law, etc.)?

Most of the social statistics allow the creation of official microdata file as long as the confidentiality of the respondents is protected. Official microdata are called Public Use microdata file (PUMF).

1.2 Legal regulations concerning the dissemination of microdata

According to which rules are microdata distributed to the scientific community?

All Statistics Canada data are collected under the Statistics Act which guarantee the confidentiality of the data. Therefore, only Statistics Canada employees have access to the raw data. In some cases, the raw data could be shared with other Federal/Provincial Departments only if the respondent provided his/her consent to share the data with the third parties. Under no circumstances the general public, including the academic researchers, have access to the raw data, also called the Master File (a microdata set). The Master file is the official Statistics Canada microdata set.

In order to allow a larger access to the data, Statistics Canada produces Public Use Microdata File (PUMF) which complies to the Statistics Act. Each record on the PUMF represent a respondent. The difference between the master file and the PUMF is that some variables are suppressed or regrouped to protect the confidentiality. PUMF are available to the public for a fee and to Universities through the DLI program. In general PUMF are produced for the social field data: income, health, education, etc. The PUMF are the official product of the program.

PUMF are available for sale to the public. Under the Data Liberation Initiative (DLI) : a partnership between the Canadian Universities and Statistics Canada, the academic researcher have access free of charge to PUMF. Other users will have to purchase the PUMF

* Jeanine Bustros is Chief of the User Support Services, Dissemination Division, at Statistics Canada and member of the DLI External Advisory Committee.

2.1 Released microdata files

Which official microdata are distributed to the scientific community (e.g. Labour Force Survey, household budget, etc.)?

As long as the files are screened for confidentiality, we can create a PUMF. Labour Force survey has no microdata files nor the Consumer Price Index. The other exception is longitudinal files. Otherwise, most of the social surveys (data collected from a household) are disseminated by means of PUMF.

2.2 Formats of released microdata

2.3 Data documentation

In which form (raw data, system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata distributed?

The files are in ASCII format which proper documentation. As much as possible we try to have either SAS or SPSS command.

2.4 Consultative services

Does the Statistical Office provide consultative services to academic data users; if it does: to which extent and in which form?

Yes through the Data Liberation Initiative. Please refer to the web site

<http://www.statcan.ca/english/Dli/contents.htm> .

2.5 Organization of the data distribution within the statistical office

Is there a central unit within the Statistical Office responsible for distributing data to academic research?

Yes the DLI.

2.6 Scope of methods for disclosure control

To which extent are measures taken to anonymize data (combining of categories, removal of variables, identifiers, etc.)?

PUMF are anonymous by definition.

3. Release of microdata to data archives and/or individual researchers

Are the data distributed to a data archive or other central scientific institution for further transfer to researchers/research institutes or only to individual researchers/research institutes upon request?

For archive, all statistical products, including PUMF are sent to Statistics Canada Library and

the National Archive Library.

4. Rights of data users

Which rights does the data recipient acquire? Can a data recipient, for example a data archive, transfer data to third parties; if this is possible: which are the conditions; or does the data recipient have to use the data exclusively for himself/herself?

Access are governed by a licence agreement. So depending on the terms, it may happen that the data is used by a third party.

5. Cost

Which are the total costs charged by the Statistical Office for the distribution of data, what are these total costs composed of (costs for data, direct costs for data distribution, e.g. for data carriers, costs for the production of photocopies, for staff, a proportion of the costs of/flat rate for the generating of data, data documentation costs, etc.)? If data are supplied to a scientific institution, for example a data archive, which costs are charged to the final user?

Costs vary depending on the program. Several PUMFs are available on CD-ROM and are sold for \$2,000 CDN each. For academic institutions (universities and colleges) which are members of the DLI, the PUMFs are free of charge. However there is an annual fee to become member of the DLI: \$3,000 for small universities and colleges, and \$12,000 for large universities. At the moment there are more than a 100 titles/product on the DLI collection.

6. Experiences of the statistical office

Which are the experiences made by the Statistical Office regarding the transfer of data to academic research? (With respect to the usability of research results, how discreet data are handled by scientists, research co-operation between the Statistical Office and academic research, etc.).

There is very little knowledge on how the data is used by academic research. As a pilot project, the DLI couldn't fulfil this need. However, we are planning to work with the universities to strength the feedback on data usage.

3B) Canada: University of Alberta, Data Library (Local DLI Contact)

Charles Humphrey *

1. Conditions of data acquisition

Under which conditions can the data archive obtain official microdata from the Statistical Office?

Prior to a relatively new subscription program with Statistics Canada called the Data Liberation Initiative (DLI), each public use microdata file (see the answer to Question 10 for a clarification of public use microdata files) was purchased from our national statistical agency and required a separate license. The license restricted those who could access the data and for what purposes the data could be used. Under no circumstances did Statistics Canada permit the redistribution of the data.

Since the DLI pilot project began in 1996, a standard institutional license was established to cover who has access to these data. This currently includes all researchers affiliated with the university subscribing to DLI, all currently enrolled students and the university's employed staff. An annual fee is paid to Statistics Canada that permits access to all of the Agency's public use microdata files in their collection. The change in licensing policy went from an individual-based license to an institutional-based license.

To summarize, the conditions consist of paying an annual subscription fee to be eligible for access to all public use files available in the Statistics Canada Catalogue. A license permits access to University members (students and staff) but only for scholarly activities, that is, the data cannot be used by staff to perform contract or privately funded research nor may the data be redistributed outside the terms of the license (that is, researchers at two DLI institutions are allowed share a file between them in a collaborative research project). The terms are defined on the DLI web site at: <http://www.statcan.ca/english/Dli/dli.htm> .

Which are the features of, e.g., the contracts concerning the transfer of data?

The contract permits member universities of DLI to share authenticated access to data across institutions. For example, my institution provides access to Census data files to other DLI universities. However, the security features of the HTTP server are invoked to filter out requests that don't come from an authorized institution. In other words, the DLI contract allows inter-institutional collaboration to develop interfaces to the Agency's microdata products. If provided over the Internet, secure measures must be taken to authenticate access. Only other

* Charles Humphrey is head of the Data Library at the University of Alberta and co-chair of the DLI External Advisory Committee.

DLI institutions may have access.

The bottom line is that DLI data may not be redistributed outside DLI institutions.

2. Acquisition concept of the data archive

Which official microdata are requested by the data archive? Which criteria is the selection of data, if necessary, based on?

DLI operates through a dedicated Statistics Canada FTP server. Each subscribing university has open access to all of the files contained on this FTP server. Furthermore, Agency microdata files that are not part of the FTP collection may be requested and when available within the Agency, are added to the FTP site. Statistics Canada also sends out specific standard products on CD-ROM. These go to all requesting DLI member institutions.

Because of immediate Internet access by DLI institutions, many institutions operate on a just-in-time acquisition model rather than just-in-case model. That is, many institutions download a file when a patron specifically asks for a microdata file. Some of the DLI institutions have a long-established Data Library, such as our University. We have a policy of collecting data files relevant to specific research interests. If a Statistics Canada microdata file falls within our collection policy, we will access the file and incorporate it as part of our local data file collection.

Which possibilities does the scientific community generally have in the respective country to use official microdata, apart from the conditions the data archive established?

If a researcher wishes to use a Statistics Canada microdata file outside the terms of the DLI license, they are free to purchase access directly from Statistics Canada and operate under that product's individual license. Most Statistics Canada microdata files cost between \$2,000 and \$3,000 CDN.

3. Data formats, documentation, consultative services

In which form (raw data, portable system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata transferred? How are the data processed by the archive, is additional documentation provided? Which consultative services does the data archive provide to academic data users with respect to official microdata, to which extent and in which form?

DLI is evolving and as it does, so are the file formats. Currently, DLI provides raw ASCII data files with machine-readable data documentation and an SPSS command file to process the raw data file. In some instances (e.g., for older microdata files from the 1970's and 1980's), only paper documentation is available.

DLI is now working on processes to establish DLI standards for its microdata collection. Statistics Canada is composed of a large number of author divisions, all of which practice their own data management standards. Consequently, DLI receives microdata products in a variety of

formats. As I mentioned in the previous paragraph, DLI currently requests from author divisions a raw data file and machine-readable data documentation. Verification checks (number of records per file and record lengths) are performed by the Unit to ensure completeness of products. The DLI Processing Unit then prepares an SPSS command file if the author division has not produced one (SAS is used primarily in Statistics Canada so not many of the divisions produce SPSS command files). Some checks are also run between the data documentation and the data to ensure the accuracy of the documentation as well as the data.

We are now promoting within DLI the establishment of the DDI* standard for data documentation. This standard, which is based on a mark-up language and usable with XML, will require DLI to convert the variety of machine-readable data documentation that is currently received into a single DDI format. This approach is now being considered by the DLI External Advisory Committee and the cost of implementing the DDI standard is being evaluated.

The DLI Unit in Statistics Canada serves as a clearing house for questions from member institutions. Each university has one person designated as the DLI Contact. This person is responsible for communicating with the DLI Unit on behalf of her or his university. If questions arise at member institutions that the DLI Contact cannot answer, the DLI Contact uses an open email discussion list to post the query to the DLI Unit and all other DLI Contacts. If a DLI Contact at another member institution knows the answer, he or she is welcome to post an answer to the discussion list. If the answer has to come for the Statistics Canada author division, the DLI Unit will pursue obtaining the answer within Statistics Canada channels. The consultation model is distributed, that is, the first point of contact is with the DLI Contact at each institution. If the answer must come from higher up the network, then the DLI Contact is responsible for requesting the information.

4. Rights concerning data usage

Can official microdata be distributed to third parties, researchers, institutes, etc.? Under which conditions?

No. The license prohibits redistribution outside the conditions of use and access. The results of all research are expected to be public information. Thus, a researcher working on a topic for a scholarly publication can share her or his results with third parties or institutes. However, the data may not be shared.

Which rights does the data recipient acquire? Can a data recipient, for example a data archive, distribute data to third parties; if this is possible: which are the conditions; or does the data

* The project Data Documentation Initiative (DDI) was initiated by the ICPSR in 1995. With support from the National Science Foundation, among others, the DDI has produced a Document Type Definition (DTD) for metadata. Beside international data archives, statistical offices from the U.S.A and Canada are members of the DDI (for further information see URL <<http://www.icpsr.umich.edu/DDI/codebook.html> >; accessed 02.19.01).

recipient have to use the data exclusively for himself/herself?

No redistribution rights are granted outside the circle of existing DLI member institutions. The use of the data is only for scholarly research. Such research cannot be funded through contract work, that is, where the scholar is paid to perform the analysis for a third party. If a researcher wishes to engage in contract research, she or he can obtain, for a fee, the microdata file that they require directly from Statistics Canada. They will then have their own exclusive license for the data with Statistics Canada. DLI access is provided to researchers who receive funding through recognized funding agencies, such as the Social Sciences and Humanities Research Council of Canada.

5. Modalities of data access

What do researchers who are interested in official microdata have to do to obtain copies of data or to use them? What is the role of the data archive in this case?

A researcher wishing to use a DLI microdata file must see her or his DLI Contact. At some DLI institutions, this is a formal process where the researcher must make an appointment with the DLI Contact. In other institutions, the request process is quite informal, that is, the researcher may make her or his request in the form of a phone call or an email message. A few DLI institutions have set up intra-net access to their local patrons using data extraction services. In these instances, the Data Library has a local campus network site that researchers can use from their desktops to select microdata files and subsets of variables and cases. The approach taken by a DLI institution is dependent upon the service model that the institution has employed. Almost all DLI Contacts in Canada are in the university Library. Many of these libraries provide reference services to support DLI microdata.

6. Acquisition cost

Which costs are charged to the data archive a) for the acquisition of data by the Statistical Office, b) for the processing, etc., of data, c) for user services?

The DLI license is a flat fee per year of \$12,000 for institutions belonging to the Canadian Association of Research Libraries (CARL) and \$3,000 for all other institutions. The 27 largest research institutions in Canada belong to CARL. It is important to note that Statistics Canada also makes an annual contribution of dollars to the operation of DLI. Currently, the university subscription fee covers around 70 percent of the costs of providing DLI services. Statistics Canada and other government agencies cover the remaining 30 percent (or \$175,000). Each university covers all local expenses for providing access to DLI. This includes the cost of the DLI Contact (who usually wears many hats and is not just responsible DLI services) and all infrastructural costs, such as computing equipment.

7. Fees charged by the data archive

How high are the costs charged by the data archive for data transfer, what are these costs composed of (costs for data, direct costs for the distribution of data, e.g. for data carriers, costs for the production of photocopies, for staff, a proportion of the costs/flat rate for the generating of data, data documentation costs, etc.)?

The cost model is rationalized on the recovery of revenue that author divisions within Statistics Canada would lose based on traditional sales by divisions to universities. Each division is compensated through the DLI subscription fee for the division's microdata contributions to the DLI project.

8. Financing

How is the archive financed? Especially: who bears the costs that emerge for the acquisition of microdata if the data archive's funds do not cover the acquisition/distribution of that data?

As mentioned above, the DLI Unit is funded by university subscription fees (27 CARL institutions paying \$12,000 per year and 38 non-CARL institutions paying \$3,000 per year) and an infusion of \$175,000 CDN by Statistics Canada and a few government departments.

Each university is responsible for their local costs in providing access to DLI data.

9. Demand for official microdata

How strong is the demand for official microdata, how extensive is the provision of consultative services?

DLI has revolutionized access to Statistics Canada microdata files within the academic community in Canada. All institutions, regardless of size or proximity to Ottawa, have equal access to microdata files. The demand for these files continues to grow almost on a monthly basis as more researchers become aware of DLI on their campuses.

While the playing field for access to data has been leveled through the DLI project, the type of consultative support is highly dependent upon the investment of each university. Many of the smaller universities cannot afford to dedicate a full-time employee to the DLI Contact position nor can they afford to train this person at the level that the Contact can field the vast majority of the questions asked. Nevertheless, DLI has undertaken a base-line training program and presented regional workshops across Canada preparing DLI Contacts with the skills they require for minimal service. Contacts have also been provided with referral support services when they cannot answer a data question.

10. Experiences of the scientific community

Which are the experiences made by academic research concerning the cooperation with the Statistical Office? Where could improvements be made from the scientific community's point of view?

In Canada, the largest improvement from the scientific community's point of view is to create some method of access to the master (confidential) data files. That is, only public use microdata files are available under DLI. A public use microdata file differs from the master file in that Statistics Canada has removed information from a copy of the master file to minimize the likelihood of disclosure. This process entails steps such as collapsing detailed response categories into a few general categories (for example, instead of 7-digit occupation categories, just 17 general categories are reported), reporting only gross geography (e.g., provinces and urban regions greater than 250,000 or 300,000 people), or removing a variable entirely from the public use file. As a consequence of needing to remove information from master files to create public use files, the only public use files are cross-sectional. In other words, no public use files are now being created for the combined panels of longitudinal surveys. Simply too much information has to be removed from a longitudinal file to meet the standards of a public use file. As a result, researchers who wish to work with longitudinal panel data from Statistics Canada are left out of the loop.

A new initiative between Statistics Canada and the Social Sciences and Humanities Research Council is addressing this concern.* The proposal is to create secured research data centres in nine or ten universities across Canada that will provide access to researchers who have had their research request to use these master files approved through a peer-review process.

* See "Final Report of the Joint Working Group of the Social Sciences and Humanities Research Council and Statistics Canada on the Advancement of Research Using Social Statistics", December 1998; URL <http://www.sshrc.ca/english/policydocs/discussion/statscanreport.pdf> ; accessed 13 May 1999.

4A) The Netherlands: Centraal Bureau voor de Statistiek (CBS; Statistics Netherlands), Voorburg

Joris Nobel *

- 1.1 Legal regulations concerning collection of data**
- 1.2 Legal regulations concerning the dissemination of microdata

According to which rules/norms are official microdata compiled, and is there a legal basis for it (statistics law, etc.)?

Since 1996 there is a legal title for the restricted release of official microdata in articles 13 and 14 of the Statistics Law(enclosed [left out to save space; B.S./F.K.]). This release is one of two formal exceptions to the general obligation of statistical confidentiality (article 11; the other exception concerns the transmission of confidential data to Eurostat). The law gives the following safeguards:

- release of microdata from business surveys with a legal response obligation remains forbidden (in practice, this means that only social surveys are released).
- microdata may be released only to research institutes mentioned in the law (article 13/2; release to other research institutes is dependent on the authorisation by the independent Central Commission for Statistics (CCS). The CCS has set out criteria to be met by organisations (not persons!) that do not qualify as a university, a research institute founded by law, a governmental planning bureau, or Eurostat. The following criteria are applied to these organisations:
 1. The organisation needs to be an independent legal personality or a part of a governmental service (national, provincial, municipal). In the case of a governmental service one needs to bear in mind that this service needs to prove the criterion mentioned under 2.
 2. the organisation has to be outside the range of an administrative authority;
 3. research is the organisation's primary goal;
 4. the organisation's publications are of a public nature. If the organisation carries out a commissioned research the result for which the microdata have been used should be published;
 5. the organisation will have an excellent reputation."

* Joris Nobel is adviser of the Director-General of the CBS and Secretary of the Central Commission for Statistics (CCS).

** In general I refer to the Dutch literature as a background and source of further information to the following answers: Citeur/Willenborg (1993); de Vries/Nobel (1999); Holvast (1999); Keller et al. (1993); Kooiman et al. (1999); Nobel (1995, 1996, 1999), and Willenborg/de Waal (1996). I furthermore refer in general to the answers from WSA, the Scientific Statistical Agency founded by NWO, the Netherlands Science Foundation. As a further note I must add that CBS is reorganising itself at present. References to organisational units refer to the present (and not the new) structure.

- the Director-General of Statistics determines the level of statistical disclosure protection to be applied before release (article 13/1; see below under question 3).
- microdata are released under a contract with the research institute (article 14; see below under question 5).

According to which rules are microdata distributed to the scientific community

Since 1994 there is a multi-annual contract between Statistics Netherlands CBS and the Netherlands Science Foundation (NWO). This contract came about after a long „cold war“ between CBS and the academic world.

- Statistics Netherlands is obliged to release at least eight new editions of microdata for its major social surveys each year. These research microdata files are standardised. Documentation of metadata is part of the service provided.
- NWO pays a lump sum of one million guilders a year. It has installed a Scientific Statistical Agency (WSA) that brokers between CBS and the research community: it represents the research community in discussions with CBS about the annual package of files and about the balance between confidentiality and user value; it releases a news letter; it helps to organise user meetings.

The contract was evaluated in 1997 and prolonged for another four years in 1998. The scope of WSA has widened by now to the release of microdata from other sources than CBS only. It has also taken an interest to other modes of access to microdata such as on site and remote access. These latter modes are of interest for access to business microdata and social microdata stemming from administrative sources instead of original surveys.

2.1 Released microdata files

Which official microdata are distributed to the scientific community (e.g. Labour Force Survey, household budget, etc.)?

For all social surveys (and some minor other surveys) on the CBS programme official microdata are released (see WSA questionnaire and website <http://wsa.gamma.rug.nl>).

Since 1998 CBS is involved in an experimental three year programme (named CeReM) to make business microdata accessible on site for academic research. This programme has been requested by the academic community in several official reports. NWO has sponsored it with a grant of over one million guilders. An extensive consultation of the business respondents community and the legal and parliamentary branches of the Ministry of Economic Affairs under which CBS resides, has preceded the programme. In the second half of this year the programme will be evaluated.

2.2 Formats of released microdata

In which form (raw data, system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata distributed?

Raw data or SPSS system files, usually.

2.3 Data documentation

2.4 Consultative services

2.5 Organization of the data distribution within the statistical office

Does the Statistical Office provide consultative services to academic data users; if it does: to which extent and in which form

CBS releases documentation of the microdata as part of the service provided. CBS staff participate in the user meetings organised by WSA. CBS staff check draft reports by researchers on statistical disclosure protection aspects. Other consultation takes place on an ad hoc basis.

There is a small network of senior subject-matter statistical management, research methodologists, and policy staff. Depending on the specific question to be solved the focal point within this network may shift.

2.6 Scope of methods for disclosure control

To which extent are measures taken to anonymize data (combining of categories, removal of variables, identifiers, etc.)?

Direct identifiers are always deleted. For combinations of indirect identifying variables internal rules and criteria for statistical disclosure avoidance have been developed by the central unit for research methodology. This unit contributes to both methodology (PRAM) and software tools (ARGUS). Implementation is a task for subject matter statistical management. Full anonymous microdata (no direct but full indirect identifiers) are available for analysis at the CBS premises.

3. Release of microdata to data archives and/or individual researchers

Are the data distributed to a data archive or other central scientific institution for further transfer to researchers/research institutes or only to individual researchers/research institutes upon request?

Data are released by CBS to the research institute/researcher. The contract is signed on behalf of the research institute, each individual researcher signs a personal confidentiality statement. Microdata are archived by CBS. WSA intermediates but has no access to microdata itself.

4. Rights of data users

Which rights does the data recipient acquire? Can a data recipient, for example a data archive, transfer data to third parties; if this is possible: which are the conditions; or does the data recipient have to use the data exclusively for himself/herself?

The contract specifying the obligations and rights of both parties is enclosed in English. [dropped to save space, see <http://wsa.gamma.rug.nl/Data/CBS.pdf> ; B.S./F.K.]. In particular, the data recipient shall:

- Not store the microdata on open networks (such as the Internet)
- Give access to the microdata only to staff that is known to CBS and has signed a personal confidentiality statement
- Not match the microdata to other datasets
- Not transfer the microdata to other parties
- Destroy the microdata and documentation after their use
- Send draft publications to CBS for inspection on the maintenance of statistical confidentiality
- Contribute to security inspections by CBS staff on their premises

5. Cost

Which are the total costs charged by the Statistical Office for the distribution of data, what are these total costs composed of (costs for data, direct costs for data distribution, e.g. for data carriers, costs for the production of photocopies, for staff, a proportion of the costs of flat rate for the generating of data, data documentation costs, etc.)? If data are supplied to a scientific institution, for example a data archive, which costs are charged to the final user?

NWO pays a yearly lump sum to CBS for its release of microdata. This sum is meant to cover the contribution by the research community in the data collection costs of CBS. NWO also covers the bureau costs of WSA (two persons). Users pay an additional fee dependent on two factors: academic status (non-academic users pay the double amount) and size of the file. The academic fee varies from 1,000 to 5,000 guilders. There is a discount for the whole package and for older microdata files. Part of these revenues flow back to CBS to cover the marginal costs of data release. The remainder is used by WSA to further improve the data structure for Dutch social research.

6. Experiences of the statistical office

Which are the experiences made by the Statistical Office regarding the transfer of data to academic research? (With respect to the usability of research results, how discreet data are handled by scientists, research co-operation between the Statistical Office and academic research, etc.).

On balance, our evaluation is quite positive. The release of microdata has become more of a routine for both CBS and its users. CBS data are widely used, often for purposes and in ways that go well beyond its own mission and possibilities. There have been no serious breaches of statistical confidentiality. Policy discussions with Academia, parliament and the press which around 1990 (that is before the current arrangements) were very demanding and hardly rewarding, by now have become superfluous and non-existent. Two points of attention must be mentioned for the future, however:

- It takes a major effort to make sure that co-operation between CBS and Academia is on the basis of equal standing and appreciation. Academic users often look at CBS as a mere provider of microdata whereas CBS has higher ambitions in adding value and producing information. Sometimes (especially non-academic policy) users produce studies with CBS microdata that compete with our own publications.
- In the future the CBS data collection and processing strategy will impede the release of microdata files for each separate social survey. To an ever increasing extent background variables will be derived from the population administration, the fiscal administrations, et cetera. Already at present researchers show an active concern about their future access to official microdata. New modes of access will have to be developed to comply with their needs.

4B) The Netherlands: Wetenschappelijk Statistisch Agentschap (WSA; Scientific Statistical Agency), Den Haag

Ron Dekker *

First we have to explain the distinction between the Scientific Statistical Agency (WSA), which is part of the Research Council for the Social Sciences, and the National Data Archive (NIWI Data Archives) which belongs to the Royal Dutch Academy for the Sciences.

The Agency's focus is on opening up data for research. This concerns two main activities:

- 1) availability
 - legal and financial contract with data-owners, like Statistics Netherlands
- 2) accessibility
 - control on the quality of the data and the documentation
 - promote transfer of knowledge, e.g. by organising user-meetings

The Data Archives hold actual archiving tasks, like preserving and distributing the data. The Agency does not take care for the distribution of the data. Data-owners can deposit their data at NIWI Data Archives (comparable to Zentral Archiv, in fact the NIWI organisation is comparable to GESIS), or become their own data-provider, that is they will take care of preserving the data and distributing the data and documentation themselves.

WSA has two financial instruments to open up data:

- Long or medium term agreement between WSA and producer:
 - CBS
 - School cohorts (primary and secondary education)
 - Data of the Social and Cultural Planning Bureau 'SCP'
- Archival projects, incidental grants to improve quality of data and documentation
 - European Values Studies, Labour market entry studies of ROA, OSA labour market panel, Telepanel by CentERdata
 - CeReM, individual firm data that are accessible on site at CBS.

Recently Statistics Netherlands started a pilot which offers the opportunity for Dutch researchers to work on firm microdata (access to these data is heavily protected by law). The Research Council stresses the importance of this facility called CeReM (Centre for Research on Economic Microdata) and has granted a significant subsidy to CBS to start up CeReM.

* Ron Dekker is a member of the WSA.

Future developments of WSA*

The WSA-medium-term plan 1998-2001 “Bronnen Benut” (Sources in Use) states the strategy for the second term of four years.

For the near future the emphasis is on:

- To extend the availability of micro data to non-CBS data
- To pay special attention to internationalisation of research
 - constructing English documentation for Dutch data and
 - exploring the possibility of disclosure of relevant European data (Eustage discussion paper 1998).
- For this the Agency co-operates with other institutes, such as the NIWI Data Archives in the Netherlands.

1. Conditions of data acquisition

Under which conditions can the data archive obtain official microdata from the Statistical Office? Which are the features of, e.g., the contracts concerning the transfer of data?

- There is an agreement between CBS and WSA (confirmed by a letter from NWO to Director-General of CBS).
- NWO pays a lump sum of 1 million guilders to CBS and CBS promises to open up surveys on persons and households (originally it was stated “at least 8 surveys per year”, in practice almost all surveys are available).
- There is an agreement between user organisation and CBS, see appendix A. [dropped to save space; B.S./F.K.; see <http://129.125.158.28/Data/CBS.pdf>].

By law the Central Commission for Statistics (CCS) has to approve on access to the data by the organisation. In the CCS-CBS law art. 13, universities and planning bureaux are mentioned explicitly (to have access).

The rules for approval are:

“To obtain microdata from surveys on persons and households by Statistics Netherlands (CBS) permission is needed from the Central Commission for Statistics (CCS). The CCS has set out criteria to be met by organisations (not persons!) that do not qualify as a university, a research institute founded by law, a governmental planning bureau, or Eurostat. The following criteria are applied to these organisations:

1. The organisation needs to be an independent legal personality or a part of a governmental service (national, provincial, municipal). In the case of a governmental service one needs to bear in mind that this service needs to prove the criterion mentioned under 2.

* The complete overview of future activities of the WSA is left out to save space (B.S./F.K.).

2. the organisation has to be outside the range of an administrative authority;
3. research is the organisation's primary goal;
4. the organisation's publications are of a public nature. If the organisation carries out a commissioned research the result for which the microdata have been used should be published;
5. the organisation will have an excellent reputation."

Some important items in the contract are:

- The supplied data shall be used only for statistical analysis
- no allowance to match the supplied data on the level of individual elements
- no passing on, or supplying of data for processing or use by third parties
- ensure that the data are properly protected, both physically and logically
- The data are protected against so-called 'spontaneous recognition' (the files could be labelled as scientific-use files, contrary to public-use files, which are fully protected).

For this CBS divides variables into three categories according to their contribution on recognition. In Dutch language this can be indicated by three Z's: zeldzaam (rare, seldsam), zichtbaar (visionable, sichtbar) and zoekbaar (searchable, findbar).

- direct identifying variables, like name, address, phone number, social security or tax number
- indirect identifications, to be subdivided into

mostly identifying	region or municipality
more identifying	gender, ethnic group, nationality, country of birth
identifying	company, profession, education (including low levels of aggregation), age, marital status, type of household
other variables	

Direct identifying variables are removed from a micro data file.

In carrying out the safety procedure the data file is tested on the existence of rare combinations on variable crossings of the type *mostly identifying x more identifying x identifying*

- Each researcher with the user organisation has to sign a secrecy statement. See appendix B.
[dropped to save space; B.S./F.K.; see <http://129.125.158.28/Data/CBS.pdf>]

2. Acquisition concept of the data archive

Which official microdata are requested by the data archive? Which criteria is the selection of data, if necessary, based on? Which possibilities does the scientific community generally have in the respective country to use official microdata, apart from the conditions the data archive established?

WSA holds almost all micro surveys on persons and households. In fact, the criteria is completeness of all CBS micro data (on persons and households).

WSA has subsidised additional archival projects to extend time series (back in time) of some

surveys that are heavily used by researchers.

Data on companies are not allowed to leave the CBS office by law. In order to furnish access to these “company data” Statistics Netherlands started a pilot in 1998: Cerem, the Centre for Research on Economic Microdata, which offers the opportunity for Dutch researchers to work on site at CBS offices in Voorburg and Heerlen on firm micro data. WSA stressed the importance of this facility and has granted a significant subsidy to CBS of 1,2 Million Dfl to start up Cerem in the next three years.

3. Data formats, documentation, consultative services

In which form (raw data, portable system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata transferred? How are the data processed by the archive, is additional documentation provided? Which consultative services does the data archive provide to academic data users with respect to official microdata, to which extent and in which form?

All data are in SPSS-format. Upon request users may receive data in SAS or ASCII with a SPSS setup program. The data and documentation are put on a CD-ROM and delivered to the researcher. If the researcher needs access to the original (not protected) data, he or she can go to one of the offices and work on site. Instead of protecting the data, the output is checked upon. CBS acts as its own archive. Additional documentation is available (in MS/Word) and goes with the data. WSA monitors the quality of the documentation.

WSA puts quite some effort in transfer of knowledge. First explicit ways, like documentation etc, but also in implicit ways by means of user meetings. A user meeting is organised at the WSA office (and sometimes at CBS or a university). First, people from CBS provide information on data processing, ideas for new questions or redesign etc. In the afternoon users (researchers) give brief presentations of their results. A report of the meeting is put on the Internet. These meetings are rather informal, but very effective in transferring knowledge.

4. Rights concerning data usage

Can official microdata be distributed to third parties, researchers, institutes, etc.? Under which conditions? Which rights does the data recipient acquire? Can a data recipient, for example a data archive, distribute data to third parties; if this is possible: which are the conditions; or does the data recipient have to use the data exclusively for himself/herself?

CBS does the distribution (acts as its own archive). The recipient is not allowed to redistribute the data.

5. Modalities of data access

What do researchers who are interested in official microdata have to do to obtain copies of data or to use them? What is the role of the data archive in this case?

They send a request (by letter, e-mail or Internet form) to WSA. Note that data are delivered to

organisations (faculties), never to individual persons. WSA acts as the portal site. Not only for CBS data, but for several data producers.

6. Acquisition cost

Which costs are charged to the data archive a) for the acquisition of data by the Statistical Office, b) for the processing, etc., of data, c) for user services?

WSA (or in fact the Research Council) pays a lump sum of 1 Million guilders per year to CBS.

7. Fees charged by the data archive

How high are the costs charged by the data archive for data transfer, what are these costs composed of (costs for data, direct costs for the distribution of data, e.g. for data carriers, costs for the production of photocopies, for staff, a proportion of the costs/flat rate for the generating of data, data documentation costs, etc.)?

There is a tariff for users, in which we distinguish between academic and non-academic use. Moreover there is a distinction between actual (in general: the three most recent years that are available) and historical data

Tariff Structure

Price rate	actual	historical
academic	100%	50%
non-academic	200%	100%

The tariffs of the surveys are in the table below. Note there is an upper level of 2.275 Euro for a survey.

Delivery costs of the data are included in the tariff. WSA and CBS negotiate on the delivery costs. Currently, CBS charges 250 Euro for each delivery (e.g. the Faculty of Sociology at Tilburg University orders the Labour Force Survey 1998. They have to pay 2275 Euro. CBS takes care of billing. They keep 250 Euro for delivery costs and pay 2025 Euro to WSA.

Tariff Micro Surveys CBS, actual data at academic tariff, price per data file (year)		
survey	description	Euro
VOCL89	Secondary Education Cohort 1989	1.365
VOCL93	Secondary Education Cohort 1993	1.365
EBB	Labour Force Survey	2.275
SEP	Socio-Economic Panel	2.275
SEP Longitudinaal	Longitudinal version of SEP 1984-1995	2.275
BO	Budget Survey	1.365
OVG	Mobility Survey	2.275
PAP	Car (PKW) Panel	455
MFE	Motor Bike Survey	455
BVE	Company Vehicle (LKW) Survey	455
CCO	Consumers' Confidence Survey	455
OG	Family Survey	910
B&U	Building and Construction Registry	455
OD	Daily Recreation Survey	1.820
OIT	Survey on Tourism	910
NKO	National Election Study	455

Tariff Micro Surveys CBS, actual data at academic tariff, price per data file (year)		
survey	description	Euro
DLO	Living Conditions Survey	1.365
WBO	Housing Needs Survey	2.275
GE	Health Survey	1.365
ESM	Survey on Crime Victimisation	910
ERV	Survey on Safety	910
TBO Old version	Time Budget Study	455
POLS	Permanent Study on Living Condition (POLS)	
BASIS	Base module POLS	455
GE-AROM	Health and Labour Conditions	1.820
REM	Law and Environment	910
REP	Law and Victims	910
TBO	Time Budget	455
JONG	Living Conditions of Youth	455
SLI	Social and Cultural Living Conditions Index	455
OiN	Accidents in the Netherlands	
WONEN	Housing	2.275
Abb.	Subscription (all actual data of one year)	11.375

8. Financing

How is the archive financed? Especially: who bears the costs that emerge for the acquisition of microdata if the data archive's funds do not cover the acquisition/distribution of that data?

The lump sum contract (450.000 Euro) with CBS is paid by the Research Council NWO (at top level).

The Social Sciences Research Council invests annually 120.000 Euro for acquisition of new data.

WSA must finance its own staff. Annually WSA receives about 150.000 Euro from delivery of data. This is sufficient to finance the WSA staff, which is kept very small: only 2.1 fte or 100.000 Euro per year, and to invest in new acquisitions.

9. Demand for official microdata

How strong is the demand for official microdata, how extensive is the provision of consultative services?

The need for micro data has been very large.

During the period 1994-1998 there were about 300 orderings. About 40 were subscriptions, which means that customers get all the data (about 12 surveys) at one delivery.

In 1999 there were 71 orderings, by 35 organisations.

About 70 organisations have ordered CBS data at least once: there are 45% university users, 15% governmental organisations (ministries, planning bureaux), 40% research institutes (outside the university). There are 8 subscriptions, which are mainly held by faculties of universities.

We can only estimate the number of individual users. In our 1997 evaluation report this was estimated to be at about 1,000 individual users.

10. Experiences of the scientific community

Which are the experiences made by academic research concerning the co-operation with the Statistical Office? Where could improvements be made from the scientific community's point of view?

The mutual understanding between CBS and research community has boomed. Access to the data has improved and is feasible at lower costs.

Improvements would be on the degree of data protection (aggregating or truncating values).

Due to privacy regulations there is low or no access to regional data.

5A) Norway: Statistisk Sentralbyrå (SSB; Statistics Norway), Oslo *

1.1 Legal regulations concerning collection of data

According to which rules/norms are official microdata compiled, and is there a legal basis for it (statistics law, etc.)?

The tasks and organisational principles of the Statistical Office are set forth in the 1989 Statistics Law. This is a framework law which establishes procedures and principles, but the programs for collecting data are not enshrined in law. Rather, the Director of the Statistical Office provides long-term plans, a budget, the working program, and a cost accounting each year to an oversight committee; that committee is then in charge of conveying this information to the relevant Ministry for adoption and passage.

1.2 Legal regulations concerning the dissemination of microdata

According to which rules are microdata distributed to the scientific community?

Researchers and planners are given particular responsibilities with regard to information, under the Statistics Law, and a data inspectorate makes decisions over the use of the statistics. The Data Privacy Law states that the Statistical Office can make microdata available for research use for a limited time, under the following conditions:

- Data must be anonymous, whereby the statistical methods used to limit disclosure are not allowed to limit their use in research.
- Non-confidential data may only be made available by the Statistical Office once the data inspectorate has given its permission to do so.
- Those who receive data are obligated to respect secrecy provisions, and violations are punishable by law.

The NSD functions as a trustee through agreements reached between the Norwegian Social Science Data Archive NSD (Norsk Samfunnsvitenskapelig Datatjeneste, Bergen), the Statistical Office, and the data inspectorate. This NSD trusteeship includes bona fide research conducted with semi-confidential data.

{ 1.3.2 Requirements concerning access to microdata:

Personal data from the filing system must not be disclosed except for research and public planning. The recipient must have the necessary licence from the Data Inspectorate. The recipient is subject to a duty of secrecy of the Statistic Act and must therefore not disclose publish data relating to an identifiable person or company. }

* Information from an earlier report (Schimpl-Neimanns/Kraus 1996) was supplemented by replies of the SSB to a KVI survey, made available to us by the Federal Statistical Office of Germany. These replies are indicated by curved brackets { }.

2.1 Released microdata files

Which official microdata are distributed to the scientific community (e.g. Labour Force Survey, household budget, etc.)?

The Statistical Office provides the NSD with a large variety of microdata (Census samples, employment samples, budgets, etc.) for archiving and for controlled dissemination of data to be used for research purposes.

{3. Microdata offered by SN:

Public use files are offered without restrictions; scientific use files are given only to qualified scientists/institutions.

3.4 Eligible for access to the scientific use files / 3.5 Restrictions:

Data for scientific use can only be given to researchers in universities and public funded research institutions. Foreign users can get data on certain conditions, one of them is that Statistics Norway has got permission from The Data Inspectorate to supply the data. / The use of the files are restricted to specified research project. }

2.1 Formats of released microdata

In which form (raw data, system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata distributed

{4.1.1.2 Format of the data:

Mostly ASCII, but also SAS and import files for SPSS. }

2.3 Data documentation

{4.1.1.4. Data documentation:

In SAS files metadata goes with the data. Otherwise the data is documented separately. }

2.4 Consultative services

Does the Statistical Office provide consultative services to academic data users; if it does: to which extent and in which form?

Each specialist branch in the Statistical Office provides user support. However, researcher support comes primarily through the NSD, with the NSD relying on the Statistical Office as needed.

2.5 Organization of the data distribution within the statistical office

Is there a central unit within the Statistical Office responsible for distributing data to academic research?

Dissemination exclusively through the NSD.

2.6 Scope of statistical methods for disclosure control

To which extent are measures taken to anonymize data (combining of categories, removal of variables, identifiers, etc.)?

Coarsening, removal of small area identifiers.

4. Release of microdata to data archives and/or individual researchers

Are the data distributed to a data archive or other central scientific institution for further transfer to researchers/research institutes or only to individual researchers/research institutes upon request?

Microdata on persons/households in anonymous form from practically all subject domains is regularly passed on to the NSD from the Statistical Office. In its trustee capacity, the NSD administers as well as disseminates this data to individual researchers.

4. Rights of data users

Which rights does the data recipient acquire? Can a data recipient, for example a data archive, transfer data to third parties; if this is possible: which are the conditions; or does the data recipient have to use the data exclusively for himself/herself?

{4.1.1.6. Signed contract for data use: Yes

4.1.1.7. Limited period of time. They have to be deleted afterwards

4.1.1.8 Use of data is restricted to specific purpose}

5. Cost

Which are the total costs charged by the Statistical Office for the distribution of data, what are these total costs composed of (costs for data, direct costs for data distribution, e.g. for data carriers, costs for the production of photocopies, for staff, a proportion of the costs of/flat rate for the generating of data, data documentation costs, etc.)? If data are supplied to a scientific institution, for example a data archive, which costs are charged to the final user?

Data acquisition through the NSD is centrally financed by the Norwegian research community. End users need only pay the NSD for the costs of data documentation material and for the data carriers.

6. Experiences of the statistical office

Which are the experiences made by the Statistical Office regarding the transfer of data to academic research? (With respect to the usability of research results, how discreet data are handled by scientists, research co-operation between the Statistical Office and academic research, etc.).

{4.1.1.12. Advantages for the users / 4.1.1.13 Demand for public and scientific use files in 1999:

The users have access to data on individuals from many different registers and as panel data. We suppose that is an advantage. / About 21 scientific or planning projects were supplied with data from Statistics Norway in 1999. Each of the data deliveries contained one or more files.}

5B) Norway: Norsk samfunnsvitenskapelig datatjeneste (NSD; Norwegian Social Science Data Services), Bergen

Vigdis Kvalheim *

1. Conditions of data acquisition

Which possibilities does the scientific community generally have in your country to use official microdata, apart from requesting data from your organisation?

General access on application to the register/institution holding the data (some times with concession (licence) from the Data Inspectorate, etc).

If such a general access exists, is it limited to certain types of data only? Survey data on persons and households; survey data on enterprises; register data (e.g., social security records, etc.)?

No

2. Acquisition concept of the data archive

Under which conditions can your organisation obtain official microdata? Which are the basic features of, e.g., the contracts concerning the transfer of data to your organisation?

There is a written agreement between Statistics Norway (SN) and NSD concerning dissemination of data from SN to the research community. The agreement includes a number of conditions with respect to delivery procedures as well as the rights and obligations for Statistics Norway and NSD. For example it states that Statistics Norway shall determine restrictions on distribution and use of any particular set of data at any time. Furthermore it states that Statistics Norway is committed to keep NSD informed about all sample surveys and to send NSD a copy of the questionnaire and give a probable delivery date as soon as it is finalised.

NSD, on the other hand, is obliged to establish procedures and instructions to be approved by Statistics Norway, for data protection, processing, storage and dissemination. NSD also have to report on any errors or deficiencies detected in the data, twice a year report on access to individual data, and finally send a copy of publications based on data from Statistics Norway.

In practice data is transferred to NSD as soon as Statistics Norway has processed the data and usually the only restriction is that data shall not be presented before Statistics Norway has published their data. NSD may also order data that requires special processing.

Which official microdata are actually requested by your organisation?

KIRUT (Norwegian abbreviation for clients entering, going through and leaving the national social security systems) is a database containing information on the movements of individuals

* Vigdis Kvalheim is head of the Department of Administration of the NSD. The Privacy Issue Unit and the

between jobs, various National Insurance benefit schemes and the social assistance system. The database now holds information on 10% of the Norwegian adult population (16-67 years old) for the years 1989 to 1996.

Data are supplied by the National Insurance Administration, the Directorate of Labour and Statistics Norway on the following categories:

1. Background information: i.e. marital status, children, spouse's status in the social security system, place of residence, citizenship, emigration, education (including continuing education)
2. Income: including income history before 1989, as well as income history of spouse
3. Wage activity: by industry and hours worked per week
4. Job-hunting and Work Training Programs
5. Sick pay by total amount
6. Rehabilitation
7. Disablement pension
8. Widow(er)s pension
9. Transitional Benefits for Single Providers
10. Social assistance

National register of clients in substance abuse treatment institutions

At the instigation of the Ministry of Health and Social Affairs, the Norwegian Directorate for the Prevention of Alcohol and Drug Problems has developed a national documentation system for substance abuse treatment projects. The system is built up of several modules, one of them being a "National register of flows of clients" by which individual client careers through the treatment system are monitored.

The Directorate has assigned the development and operation of the register to the NSD. The project includes collecting, registering and archiving of these data. We will also supply statistical data and graphics for publishing on the basis of the client register and a register holding administrative data on the treatment institutions.

The register of flows of clients holds data on gender, age, admittance and discharge from institutions, type of institution, and type of treatment. Encrypted national identity numbers are used as client pseudonyms and individuals cannot be identified.

Which criteria is the selection of data, if necessary, based on?

Requests from researchers (users), research programmes, etc.

3. Data formats, documentation, consultative services

In which form (raw data, portable system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata transferred to your organisation?

SPSS, SAS, Raw Data

How are the data processed by your organisation, is additional documentation provided?

SURVEY DATA: Presently documents as SPSS-files, Future: NESSTAR/DDI, xml-files.

REGISTER DATA: KIRUT: raw data into relational database; National register of clients in substance abuse treatment institutions: paper, raw data (ASCII) and spreadsheet into relational database.

All documentation: Hand made at the NSD and presented at the WWW.

Which consultative services does your organisation provide to academic data users with respect to official microdata, to which extent and in which form?

What ever requested. Oral, on paper and on Internet. Documentation, sampling, data set design, software to use, processing, programming, accessing etc, etc

4. Rights concerning data usage

In which ways can the official microdata transferred to your organisation be accessed by third parties and which are the major conditions for access?

Survey-data: As anonymous files after application to NSD and after signing a pledge of secrecy. Factual anonymous files after application to NSD and the Data Inspectorate, licence from the Data Inspectorate and after signing a pledge of secrecy.

5. Modes of data access

What do researchers interested in official microdata have to do to obtain copies of data or to use them via other modes of access (on-line access, secure area research?) What is the role of your organisation in these cases, and how do you ensure feedback between data users and data collectors?

Fill in an application and send it to the NSD, apply for concession (licence from the Data Inspectorate) via the NSD, sign on pledge of secrecy etc.

Which are the major rights and obligations of the data users?

Users are required to submit a written application to NSD and to sign a pledge of confidentiality to gain access to individual data. Access is given to a specific person for a project described in detail. In order for a student to gain access to data, he/she must submit an Advisor Declaration. Thus, the advisor is to ensure that data is correctly used.

Data is obtained for personal use only.

Applicants gaining access to data are required to make reference to Statistics Norway as the source in any publications and/or tables. The reference is also to state that NSD was the

distributor of data. They are also required to send NSD three copies of any publication, which is entirely or in part based on data from Statistics Norway distributed by NSD.

Any publication based on the data should be sent to NSD in three copies. NSD will send a copy to the data register, e.g. Statistics Norway.

6. Acquisition cost

Which costs does the data provider (statistical office, register authority) charge to your organisation

a) For the purpose of partial refunding of data collection costs?

We do not pay for the purpose of refunding of data collection costs.

b) for the processing, etc., of data (anonymization etc.)?

The agreed price is lower than the marginal cost which is common for services from Statistics Norway.

c) for user services?

-

7. Fees charged by the data archive

How high are the costs charged by your organisation to a data user and what are these costs composed of (refunding of data acquisition costs; refunding of your organisation's costs for data processing; costs for the distribution of data, e.g. for data carriers; data documentation costs, refunding of staff costs)?

Access to data is free of charge for the academic users.

If we provide extra data services like questionnaire design, data analyses, etc., we charge to refund staff costs.

8. Financing

How is your organisation financed? Especially: who bears the costs that arise for the acquisition of microdata if the data archive's funds do not cover the acquisition/distribution costs of these data?

Special grants from the Research Council of Norway. Special grants from respective Ministries.
The general funding of the archive.

9. Demand for official microdata

How strong is the demand for official microdata, how extensive is the provision of consultative services?

In 1999 NSD serviced 160 projects with one or several datasets (SNs survey data).

In 1998 NSD serviced 65 projects, 57 researchers with data from KIRUT. This involves approx. 250-300 data deliveries, as this projects ordering new data over a two-three year period.

How is steering of your data services organised? (External boards, relation to Research Council)

NSD is an organisation under the Research Council of Norway, i.e. we are employed by the RC and responsible to the RC. NSD is governed by a Board appointed by the Research Council

10. Experiences of the scientific community

Which are the experiences made by academic research concerning the co-operation with the Statistical Office? Where could improvements be made from the scientific community's point of view?

There could always be improvements. There could be more resources in Statistics Norway as well as at NSD to prepare new data for research purposes (build time-series, link data from different register, etc.)

In general, both Statistics Norway and the academic community have benefited from the co-operation between NSD and Statistics Norway. By relying on NSD as a data broker to the research community, Statistics Norway is relieved from greater parts of its obligations and services to the academic community. It also benefits from data being used and the quality assurance of the data by researchers. This consequently adds to its legitimacy. Important in this respect is the fact that the Norwegian research community is very satisfied with the service Statistics Norway provides through NSD.

For the academic community the agreement between NSD and Statistics Norway, is without question the single most important factor facilitating social empirical research in Norway. By using NSD as a gateway to resources, procedures have been established (ethical, institutional and technical) that:

- Allow for external access to micro data (including very sensitive linked micro-data) in such a way that the identification of units is prevented.
- Allow for access to data, which cannot be anonymized on a factual level.
- Ensure the continuity of data deliveries.
- Ensure easy and equal (free) access to data.
- Improve the quality of documentation and data.
- Solve (to a great extent) problems involving the costs of factual anonymization and services.

6A) Austria: Statistik Österreich (Statistics Austria), Wien

Peter Findl *

1.1 Legal regulations concerning collection of data

According to which rules/norms are official microdata compiled, and is there a legal basis for it (statistics law, etc.)?

The Federal Statistics Law of 2000 is the framework law (see our homepage at www.oestat.gv.at). Based on Enclosure 1 of this law (which lists the objects of investigation) decrees are issued which regulate the individual surveys. Also laws, which in addition to substantive regulations also order statistical surveys (for example, the University Studies Law, the Goods Conveyance Law) and laws which only regulate statistical surveys (e.g., Census Law, Cancer Statistics Law).

1.2 Legal regulations concerning the dissemination of microdata

According to which rules are microdata distributed to the scientific community?

Par. 31 of the Federal Statistics Law of 2000 states that data containing no references to persons can be provided to professionals and research institutions once an agreement has been reached about the concrete use of this data within a research project, as well as over appropriate cost compensation. Care must be taken to ensure the impossibility of identifying personal data and of storing such data on external data carriers.

2.1 Released microdata files

Which official microdata are distributed to the scientific community (e.g. Labour Force Survey, household budget, etc.)?

1991 Building and Apartment Census
1991 Census
1991 Workplace Census
Microcensus; Basic and Special Surveys
Industry; Economic Surveys
Industry; 1995 Structural Surveys
Consumer Surveys
University Surveys
Community Financial Reports
Schools
Vital Records
Building Registry
Tourism
Foreign Trade

* Dr. Peter Findl is Secretary-General of Statistics Austria and head of the Information Management Department.

Motor Vehicles
Civil Aviation
Danube Goods Traffic
Road Goods Traffic
Traffic Accidents

In the case of information that is provided to Ministries and to state governments, it is not possible to state what the further dissemination is to various institutes (using this information for research).

2.3 Formats of released microdata

In which form (raw data, system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata distributed?

Usually raw data (with PC formats: ASCII without delimiter, sometimes with delimiter).

2.3 Data documentation

[see Appendix 6B]

2.4 Consultative services

Does the Statistical Office provide consultative services to academic data users; if it does: to which extent and in which form?

Yes, but until now at no cost (this will change and fees will be levied).

2.5 Organization of the data distribution within the statistical office

Is there a central unit within the Statistical Office responsible for distributing data to academic research?

No

2.6 Scope of methods for disclosure control

To which extent are measures taken to anonymize data (combining of categories, removal of variables, identifiers, etc.)?

Usually this has only been a removal of names and addresses, if they were even collected, as well as information about the region the community lies in. District-level information, however, remains available.

3. Release of microdata to data archives and/or individual researchers

Are the data distributed to a data archive or other central scientific institution for further transfer to researchers/research institutes or only to individual researchers/research institutes upon request?

Both (central function: WISDOM)

4. Rights of data users

Which rights does the data recipient acquire? Can a data recipient, for example a data archive, transfer data to third parties; if this is possible: which are the conditions; or does the data recipient have to use the data exclusively for himself/herself?

Only exclusive use is permissible. Disseminating the data further or even merely reusing it for another project within the same research institution necessitates renewed approval by Statistics Austria.

5. Cost

Which are the total costs charged by the Statistical Office for the distribution of data, what are these total costs composed of (costs for data, direct costs for data distribution, e.g. for data carriers, costs for the production of photocopies, for staff, a proportion of the costs of/flat rate for the generating of data, data documentation costs, etc.)? If data are supplied to a scientific institution, for example a data archive, which costs are charged to the final user?

Up to this point in time, costs incurred as a result of dissemination were passed on, which is to say that no costs for the data collection itself were charged. No commissions for dissemination were calculated (as long as it was for research purposes and not for a profit-oriented enterprise). As Statistics Austria has been made independent of the Federal administration since 1 January 2000, the cost calculation guidelines will be revised.

6. Experiences of the statistical office

Which are the experiences made by the Statistical Office regarding the transfer of data to academic research? (With respect to the usability of research results, how discreet data are handled by scientists, research co-operation between the Statistical Office and academic research, etc.).

No negative experiences; rather some positive ones which have come about through informal co-operation.

6B) Austria: Wiener Institut für Sozialwissenschaftliche Dokumentation und Methodik (WISDOM; Vienna Institut for Social Science Documentation and Methods), Wien

Günther Nemeth *

1. Conditions of data acquisition

Under which conditions can the data archive obtain official microdata from the Statistical Office? Which are the features of, e.g., the contracts concerning the transfer of data?

There are no official forms to complete to order microdata (Microcensuses) from Statistics Austria; one need only request them, though WISDOM can only pass these data on with the approval of Statistics Austria. WISDOM covers the normal costs of transferring these data, and there are no special conditions attached, nor are there contracts which explicitly regulate dissemination. The cost of a Microcensus is currently 950 Euro.

2. Acquisition concept of the data archive

Which official microdata are requested by the data archive? Which criteria is the selection of data, if necessary, based on?

WISDOM orders all available Microcensuses irrespective of whether orders have been placed for them by a researcher or an institution. We only order other microdata from Statistics Austria when requested to do so.

Which possibilities does the scientific community generally have in the respective country to use official microdata, apart from the conditions the data archive established?

As is true of all researchers, WISDOM can ask for microdata from Statistics Austria, and normally all requests are honoured against reimbursement. Generally speaking, all data requests must be approved by Statistics Austria.

3. Data formats, documentation, consultative services

In which form (raw data, portable system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata transferred? How are the data processed by the archive, is additional documentation provided? Which consultative services does the data archive provide to academic data users with respect to official microdata, to which extent and in which form?

The preparation of official raw data takes place in the archive; thus we are able to disseminate all Microcensuses in the form of SPSS data files. The data can also be provided in the form of MS-ACCESS databanks, if so requested. Documentation material usually consists of the original questionnaire and the interviewer instruction guide; these may also be downloaded

* Günther Nemeth is the director of the WISDOM.

from the Internet. Other documents, such as housekeeping books, are only available from us on request, but are available in copied form accompanying an order for data. Microcensus data are available in two forms from Statistics Austria: as raw data and in a dataset in which missing values have been substituted and data been checked for inconsistencies by Statistics Austria. WISDOM always uses the raw datasets, though in particular cases we also order an already edited dataset. We provide other documentation materials only when WISDOM needs to undertake marginal error corrections or conduct data editing. WISDOM provides services, including Microcensus content and methodological support, and in particular we provide annual aggregate data from the basic programs. We also provide assistance in the combining of individual-level microdata; for example, by dropping one-eighth of the sample, consecutive quarterly surveys can be combined. For smaller research projects, which are satisfied just with an analysis of key features, we do the calculations ourselves.

4. Rights concerning data usage

Can official microdata be distributed to third parties, researchers, institutes, etc.? Under which conditions?

WISDOM is empowered to make microdata available to third parties as long as Statistics Austria approves; up to this point, however, there has only been one case in which the dissemination of data to a commercial enterprise has been refused.

Which rights does the data recipient acquire?

Approval of the use of data is always project-specific. If the data are to be used for another project, fresh application must be made to Statistics Austria for permission. There is an exception, and that is for university instruction; permission is granted for use in a particular course and is not bound to topic.

Can a data recipient, for example a data archive, distribute data to third parties; if this is possible: which are the conditions; or does the data recipient have to use the data exclusively for himself/herself?

Those who receive data may only exclusively use it.

5. Modalities of data access

What do researchers who are interested in official microdata have to do to obtain copies of data or to use them? What is the role of the data archive in this case?

For a researcher, there are two means to obtain the data. Either they can be requested directly from Statistics Austria or be requested from WISDOM. In the first case, WISDOM is not involved unless Statistics Austria asks us to transfer the data (which can happen if the technical division at Statistics is too busy). In the case of a direct request to WISDOM, we ask for a precise yet concise project description (about one page long), which we then provide to Statistics

Austria as part of the request for approval. When this request has been granted, we send the researcher the dataset and documentation on CD. The major difference between the two ways of obtaining microdata is the form the data comes in and the delivery time. Statistics Austria only provides microdata in the form of ASCII raw data files (and for older data only in binary form), with a 2-3 month waiting period for delivery. WISDOM, by contrast, provides data in SPSS format and - as long as permission has been granted - can attend to the request the same day. Where approval has not yet been granted, we usually have to wait two weeks.

6. Acquisition cost

Which costs are charged to the data archive a) for the acquisition of data by the Statistical Office, b) for the processing, etc., of data, c) for user services?

WISDOM pays a normal user fee of about 950 Euro to Statistics Austria for a Microcensus. This sum is primarily calculated based on the quantity of data transferred. I estimate the preparation costs at WISDOM for a “normal” Microcensus at about 500 Euro (for a Microcensus of average size with a novel questionnaire). For “large” Microcensuses, as in time budget surveys, these costs can easily rise to 1,500 Euro. For Microcensuses that are repeated every year with identical wording, the preparation costs sink to around 200 Euro. In this context, one should note the following: we only receive Microcensus raw data from Statistics Austria, which in general reflects the needs of our users. The Microcensus would also be available in an edited version used as the basis for Statistics Austria publications. These were generally comprehensible to us, but researchers want to work with their own data cleaning procedures. Our efforts in data preparation consist in correctly converting the data from the form used by the Statistical Office into SPSS-readable format, the complete labelling of all variables, and a first data check. The costs of user support are hard to estimate. I would estimate that mailing costs and fulfilling the data request would come to about 50 Euro, since the documents themselves are generally available in electronic form and can be obtained through the Internet. For advice and consultation, one could only estimate that at an hourly rate (say 30-40 Euro).

7. Fees charged by the data archive

How high are the costs charged by the data archive for data transfer, what are these costs composed of (costs for data, direct costs for the distribution of data, e.g. for data carriers, costs for the production of photocopies, for staff, a proportion of the costs/flat rate for the generating of data, data documentation costs, etc.)?

WISDOM asks users to pay the same fees that Statistics Austria levies, which is to say that the Microcensuses have become more expensive over the last few years. Our charges are set at a flat rate; older Microcensuses are thus somewhat cheaper to acquire since the rate is what WISDOM paid at the time. Students can acquire Microcensus at half price; for the rates, see

the Internet. It is also possible to order just the so-called “basic program” of the Microcensus, and that is half as expensive as the full program.

8. Financing

How is the archive financed? Especially: who bears the costs that emerge for the acquisition of microdata if the data archive's funds do not cover the acquisition/distribution of that data?

The Federal Ministry for Research provides WISDOM with a yearly budget of 40,000 Euro. The problem of insufficient cost coverage thus exists not for the Microcensuses but for all the other archived studies, because the costs and effort in these latter cases is considerably higher than for the archiving of microdata.

9. Demand for official microdata

How strong is the demand for official microdata, how extensive is the provision of consultative services?

In 1999 we made about 50 Microcensuses available. 12 Microcensuses were ordered (or already distributed) from WISDOM by the end of February 2000. Requests for consultative support tend to be few since most researchers know which Microcensus they want, itself a function of the documentation already available on the Internet. Requests generally reach us via email. WISDOM can also announce the approval granted by Statistics Austria by email, but must then subsequently provide a written request, with an original signature, through the regular mails. This procedure has shown itself to be extremely efficient, and it saves time. The effort needed to provide user support, which is to say providing answers to questions about data already provided, is quite variable. As can be expected, the needs are greatest in cases where researchers are working with Microcensus data for the first time. But most Microcensus questions can be answered on the same day via email.

10. Experiences of the scientific community

Which are the experiences made by academic research concerning the co-operation with the Statistical Office? Where could improvements be made from the scientific community's point of view?

There are no results from a user survey. The data archive is quite satisfied in its work with Statistics Austria: researcher requests have been quickly processed and even when there have been tricky data problems, we have been quickly and comprehensively aided. However, the length of time it takes for Statistics Austria to provide data could certainly be shortened; having to wait for up to 3 months for a request seems too long. It would also be desirable to have datasets be prepared more quickly, or be made available sooner. It now takes 18 months after the time a survey is conducted for a Microcensus to be made available, though in the case of the “basic program” it is somewhat quicker - only 6-7 months after the survey. This also does not mean

that all researchers also receive permission to use the data through WISDOM.

Finally, the question of costs: for some users, the data costs remain too high. In the last 6 years, the cost of acquiring a Microcensus including the supplement have increased by about 140% (from 400 to 950 Euro). Now that Statistics Austria (formerly ÖSTAT) is no longer part of the Federal administration, it is likely that researchers will have to face still higher costs. Actual numbers, or information about the new rate structure, are for the moment still unavailable.

7A) U.S.A.: U.S. Bureau of the Census and other federal statistical agencies*

(e.g. Bureau of Labor Statistics, National Center for Education Statistics, National Center for Health Statistics)

1.1 Legal regulations concerning collection of data

In the decentrally organized statistical system in the USA, various official agencies and federal departments have their own offices which conduct their own statistical investigations based on independent legal mandates. The Statistical Policy Office of the Office of Management and Budget is responsible for setting the framework and for planning statistical work, as well as ensuring the cooperation of the various components of the system. The U.S. Constitution mandates a decennial Census. The legal basis for statistical surveys, as well as the working methods of the various statistics offices are anchored in part in the Freedom of Information Act (1966) which obligates all federal offices to make data public. Exceptions include surveys where special laws explicitly preclude passing on personal data - as in the cases of Census data that are only accessible to employees of the Bureau of the Census. The legal basis that is established for individual surveys, as well as the Privacy Act (1974), also define data protection criteria.

Most Census surveys fall under Title 13 of the U.S. Code, which empowers the Bureau of the Census to conduct surveys, including for other agencies, and specific regulations govern statistical investigations conducted by other federal bureaus (for example by the Bureau of Labor Statistics, or by the Social Security Administration). If the Bureau of the Census conducts surveys for other federal institutions, their particular regulations are applied (Cecil 1993).

(For examples, see Mugge (1993). 13 USC 182 also applies to the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP)).

1.2 Legal regulations concerning the dissemination of microdata

The basic law is the 1974 Privacy Act, and this is to be applied if there are no specific regulations stipulated for the dissemination of data. The Freedom of Information Act is also fundamental, as under its provisions - with a few exceptions noted below - all government information or documents can be examined.

For statistical surveys of the Census Bureau that are conducted under Title 13, the particular and strict rules for maintaining secrecy (at 13 USC 9 and 13 USC 214) state that no personal data can be disseminated. However, it is possible for researchers to be granted access to personal data if they are designated as “special sworn employees,” made legally equivalent to

* Own investigation results.

Census Bureau employees, and obligated to maintain the privacy of the data; similar rules exist for other governmental institutions that gather comparable data (Jabine 1993). Following Title 5 of the US Code, access to anonymous individual-level data for statistical research purposes can be granted once researchers sign a statement of obligation. If it can be guaranteed that no reidentification of respondents is possible in the anonymous datasets, the microdata can be made available as Public Use Microdata Files (PUMF) (for the specific criteria, see the Statistical Policy Office / Office of Information and Regulatory Affairs in the Office of Management and Budget, 1999: Checklist on Disclosure Potential of Proposed Data Releases. Prepared by Interagency Confidentiality and Data Access Group. An Interest Group of the Federal Committee on Statistical Methodology. Washington. See http://www.fcsm.gov/docs/checklist_799.doc). PUMF data are usually made available at a geographic aggregation level of regional units with 100,000 inhabitants.

Beyond making PUMF available, there are utilization possibilities for a series of datasets, though with contractual limitations that may specify purpose, limitation of length of use or persons granted user access, data file-specific confidentiality measures, data protection obligations, penalties for violation of contract, or other limitations (Duncan et al. 1993; Jabine 1993).

In case the microdata made available by the National Center for Education Statistics (NCES) through PUMF or “encrypted CD-ROM” do not meet the evaluation purposes of the researchers, the NCES can provide more information-rich data for the researcher to analyze under the following conditions: data privacy obligation and equivalent status to NCES employees, adhering to particular technical data privacy measures, providing a copy of the publications that utilized the data to the NCES, and connecting the purpose of the analysis to particular evaluation goals (Duncan 1993: 168).

2.1 Released microdata files

The statistical offices provide a multiplicity of Public Use Microdata Files which can either be provided via data carriers or can often be accessed through the World Wide Web.

The following Census Bureau and Bureau of Labor Statistics data are accessible directly from the Census Bureau through various interactive Internet programs on the World Wide Web:

The NCES provides a variety of surveys in Public Use File form, both on CD-ROM and in on-line versions that can be evaluated on the Internet (see <http://nces.ed.gov/das/htm/surveys.htm> ; <http://nces.ed.gov/pubsearch/licenses.asp>). Access to complete raw data files is only possible if a “NCES restricted data license“ is granted.

The Social Security Administration provides the following microdata:

- New Beneficiary Data (NBDS; 1982, 1991; the 1982 survey was augmented by information from register and administrative data). Available at a cost of \$150 each. Four NBDS Public Use Administrative Data files (on Retired beneficiaries, Disabled beneficiaries, Other Aged beneficiaries, Medicare Only beneficiaries) are accessible free over the Internet, in ASCII form. See <ftp://ftp.ssa.gov/pub/statistics/nbds/admin> .

Data from the U.S. Bureau of Labor Statistics (BLS) and the U.S. Department of Labor are available over the WWW from the Center for Human Resource Research (Ohio State University) on CD-ROM:

- National Longitudinal Surveys (various files, each costing about \$10-30; <http://www.chrr.ohio-state.edu/nls.html> ; for microdata with differentiated regional information, registration and special data protection declarations are necessary; see <http://stats.bls.gov/nlshome.htm>).

Following the Freedom of Information Act, official microdata can be acquired from the National Archives and Records Administration through its Center for Electronic Records. These are provided in the form of Public Use Microdata Files at low cost (maximum copying costs: \$90), and if the requested data can not be passed on for data protection reasons, the Archive can create appropriate Public Use Files (see <http://www.nara.gov/nara/electronic/>).

The USA has a well-established tradition of making public data available to researchers and to the public at low or moderate cost. At social science data archives such as the Inter-university Consortium for Political and Social Research (ICPSR; University of Michigan), for example, the following official microdata are accessible in raw data form, including setups for reading them and with standard statistical software packages (see <http://www.icpsr.umich.edu/archive1.html>):

- Census data in various versions (subpopulations and subsamples): 1790-1960 Censuses; 1970, 1980, 1990 Census
- American Housing Survey Series since 1973
- Current Population Survey Series since the 1960s (including March Individual-Level Extracts, 1968-1992; Uniform March Files, 1964-1988; Uniform October Files, 1968-1990)
- Other Census (including the 1976 Survey of Income and Education and the Survey of Income and Program Participation (SIPP) since 1984)
- Continuing Series of Consumer Surveys (since 1960/61, including Integrated Diary and Interview Survey Data, 1984-1996; 1980-1989: Interview Surveys for Household-Level Analysis)
- Various files from the U.S. Department of Education/National Center for Education Statistics (including the 1980 and 1984 High School and Beyond; National Education Longitudinal Study)
- Various files from the National Center for Health Statistics and the Social Security Administration (Health and Nutrition Examination Survey; Health Interview Survey; New Beneficiary Data System)
- American Housing Surveys (AHS) since 1980
- Survey of Income and Program Participation (SIPP) since 1984
- Survey of Program Dynamics (SPD), 1997: Experimental File [Bridge Survey]

- National Longitudinal Survey (NLS) of College Graduates, 1967-1985 (<http://www.icpsr.umich.edu/>)

2.4 Formats of released microdata

Since the early 1960s, and as a matter of course, the statistical office in the Bureau of the Census has made Public Use Files of all demographic surveys available.* The raw data of the Public Use Files are available on magnetic tape, CD-ROM, and increasingly over the last years, over the Internet, either in complete form or as a data extract.

2.3 Data documentation

Of late, comprehensive documentation about individual surveys has been available on the Internet (e.g., for the CPS, see <http://www.bls.census.gov/cps/ads/smethdoc.htm> ; <http://www.bls.census.gov/cps/pub/pubtopic.htm> ; on the Survey of Income and Program Participation, see <http://www.sipp.census.gov/sipp/>). A large-scale project to create a statistical meta-databank is under development (Lestina et al. 1996; Wallace et al. 1999).

2.4 Consultative services

Through the specialized divisions in the statistical offices. For questions about the Census, CPS and SIPP, email can be sent to the statistics offices.

2.5 Organization of the data distribution within the statistical office

Contact persons for the respective data can be found in the specialized divisions responsible for that data (depending upon the form the dissemination takes; in the Internet, entire files or extracts can be downloaded); data dissemination is partly delegated to research institutions (see point 2.1).

2.6 Scope of methods for disclosure control

At the very least, all direct identifiers (name, address, etc.) are removed in Public Use Files. Socio-demographic microdata only contain regional identifiers for at least 100,000 persons

* On the description of the beginnings of the release of Public Use Microdata Files: "In 1962, the Social Security Administration's Office of Research and Statistics began releasing microdata files on tape from their Continuous Work History Sample to other Federal and State agencies. There were essentially no restrictions on these files, and they were later used extensively by non-government researchers. The first broad release of a public-use microdata file occurred in 1963 when the Census Bureau released a file consisting of a 1 in 1,000 sample from the 1960 Census of Population and Housing. A few years later, the Census Bureau publicly released a microdata file from the Current Population Survey. Currently, unrestricted microdata files are standard products of all Census Bureau demographic surveys. They are available to any purchaser, and researchers use them extensively (Greenberg and Zayatz, 1991). Several other Federal agencies including the National Center for Education Statistics, National Center for Health Statistics, Energy Information Administration, and Internal Revenue Service currently release microdata

within a regional unit. In addition, groupings or gross generalizations and topcoding (for example in information about income) are also used. Additional statistical methods to limit disclosure depend upon the assessed risk of reidentification of the respective data (see discussions in the Subcommittee on Disclosure Limitation Methodology 1994; Zayatz et al. 1999; for a discussion of the methods for limiting disclosure with restricted data access, see Jabine 1993).

3. Release of microdata to data archives and/or individual researchers

Other than from the statistical offices, Public Use Files are also available from research data archives (like the ICPSR) as well as from commercial information providers (like the Unicon Research Corp.; see <http://www.unicon.com/>).

Particular data are only accessible under special conditions and after users have registered. For example, the Health and Retirement Study, in which income and expenditure information was provided by the Social Security Administration (with the approval of those interviewed), can only be drawn from the Institute for Social Research in Ann Arbor and can only be used under restrictive guidelines (see <http://hrs.isr.umich.edu:1041/rdapkg.htm#req0>).

4. Rights of data users

Public Use Microdata files have no limitations as to their use. By contrast, microdata that have been prepared under specific contractual regulations can only be used for specified projects and can not be passed on to third parties (see above).

5. Cost

Free download from the Internet, or at low fees for copying costs.

6. Experiences of the statistical office

There is much cooperation with the research community, as can be seen by the fact that the Bureau of the Census and the Bureau of Labor Statistics are members of the Association of Public Data Users (APDU), an association that includes both data producers and data users (see <http://apdu.org/index.html>). The Census Bureau also participates in the ICPSR “Data Documentation Initiative” project in which new documentation standards for microdata are being developed. It has also been possible for quite some time for a researcher to evaluate personal microdata while a guest of a statistics agency in the context of the “ASA/NSF/BOC Research Fellow Program” (<http://www.census.gov/srd/www/fellweb.html>). There is also close cooperation between research and statistics offices in the financing provided to Research

files” (Subcommittee on Disclosure Limitation Methodology (1994: 46).

Data Centers. Various surveys are also conducted in common or co-financed, including the Panel Study of Income Dynamics that is accessible through the ICPSR (see <http://www.isr.umich.edu/src/psid/overview.html>).

The ICPSR obtained PUMF microdata of the 1990 Census from the Census Bureau, processed and prepared them, and created a comprehensive electronic data documentation (Austin/Rockwell 1997). Another example is the cooperation between the NCES and the ICPSR in creating an international data archive on educational data at the ICPSR (1999).

What is a matter of course on the part of the statistical offices in providing research access to microdata is matched by political support for the official statistics agencies on the part of the research community. Thus, in the contentious issue of the proper statistical means of addressing a claimed undercount in the 2000 Census, and in support of scientific modes of work independent of political demands, the American Sociological Association stated that “(...) we are reaching out across research, academic, business, and public interest communities to help ensure a Census consonant with the best of science” (Levine 1996: 2).

**7B) U.S.A.: Inter-university Consortium for Political and Social Research (ICPSR),
Institute for Social Research, University of Michigan, Ann Arbor**

Erik W. Austin *

1. Conditions of data acquisition

Under which conditions can the data archive obtain official microdata from the Statistical Office? Which are the features of, e.g., the contracts concerning the transfer of data?

The U.S. has numerous official statistical agencies that produce (among other things) microdata. Most of these data are available to the public without restriction, and can be obtained for a small cost of reproduction. Increasingly, such data are being put onto the agencies' websites for free download. Most of the agencies also produce microdata that are sensitive (i.e., data in which reidentification of respondents is possible). Access is provided to these data through licenses, bundled in software that masks actual data values, or through personal researcher visits to agency-sponsored secure data "enclaves". Unlike "public use microdata files", such sensitive files are provided to a single individual who promises (in a license agreement or other such arrangement) not to redistribute the data and who agrees not to identify individual respondents.

2. Acquisition concept of the data archive

Which official microdata are requested by the data archive? Which criteria is the selection of data, if necessary, based on? Which possibilities does the scientific community generally have in the respective country to use official microdata, apart from the conditions the data archive established?

ICPSR obtains "official" microdata from statistical agencies that are typically in the public domain, and thus can be redisseminated by ICPSR. We select microdata to acquire based on our assessment of scholarly demand for such data; thus we acquire the microdata files that scholars tell us they most want to use. While U.S. scholars can also obtain these data directly from the statistical agencies, they usually prefer to get the data from us, for reasons of lower cost, greater reliability of files and technical documentation, and ICPSR's provision of user support.

3. Data formats, documentation, consultative services

In which form (raw data, portable system files for standard statistics software such as SPSS, SAS, size of the documentation) are microdata transferred? How are the data processed by the archive, is additional documentation provided? Which consultative services does the data

* Erik W. Austin is director of the Archival Development Department at the ICPSR and executive director of the Social Science History Association (SSHA).

archive provide to academic data users with respect to official microdata, to which extent and in which form?

Most of the official microdata are available from the producing agencies in raw data form, typically (even now!) accompanied by technical documentation in paper form. ICPSR processes these files by comparing data and technical documentation and correcting discrepancies, producing SPSS or SAS data definition statements, and scanning the technical documentation into electronic form, for both distribution and long-term preservation. We provide consultation on how to use these data as part of ICPSR's member services. Most such consultations now occur via e-mail correspondence, although we have a substantial telephone inquiry traffic as well.

4. Rights concerning data usage

Can official microdata be distributed to third parties, researchers, institutes, etc.? Under which conditions? Which rights does the data recipient acquire? Can a data recipient, for example a data archive, distribute data to third parties; if this is possible: which are the conditions; or does the data recipient have to use the data exclusively for himself/herself?

In general, there are no rules prohibiting redissemination of official public-use microdata in the U.S. (Thus, ICPSR is permitted to acquire and redistribute such data, and in fact usually distributes more copies of such data than does the producing agency.) Because of the problems associated with scrambled data and documentation when data are passed around indiscriminately, ICPSR does not permit redissemination of files that it supplies. This is chiefly to assure that researchers use valid, verifiable data files that we guarantee and support. ICPSR is not the only organization in the U.S. that acquires official microdata and then redistributes them; the private sector here does this on a large scale, and makes tidy profits on the sale of such data.

5. Modalities of data access

What do researchers who are interested in official microdata have to do to obtain copies of data or to use them? What is the role of the data archive in this case?

U.S. researchers interested in accessing official microdata have several choices. They can obtain such data directly from the producing agencies, usually paying a relatively low fee for a copy of a file. They can buy such data from private sector data vendors, typically for a substantial fee, and usually in a form that has been augmented by the vendor (e.g., bundled in proprietary software, merged with other information, or otherwise "improved" by the vendor). Or they can acquire data from an archive (such as ICPSR, but also including university archives and the U.S. National Archives and Records Administration). ICPSR's role in such activity is to provide access to such data at lower (or no) cost and in a more reliable manner than is often provided by the producing statistical agency. We also preserve the data in perpetuity, thus

ensuring that researchers will have access to the data years or even decades from now. Official statistical agencies rarely provide this kind of long-term access, since they are “driven” by pressures to be “current” (chiefly through their funding mandates, which are almost always presentist in nature).

6. Acquisition cost

Which costs are charged to the data archive a) for the acquisition of data by the Statistical Office, b) for the processing, etc., of data, c) for user services?

Costs charged by statistical agencies for copies of microdata files are modest: US\$ 100-200 per file is quite typical. As mentioned above, some agencies have posted such data on websites for free downloading. Most of the statistical agencies don't undertake further processing of their data, even for a fee, and the user services they offer are usually nonexistent.

7. Fees charged by the data archive

How high are the costs charged by the data archive for data transfer, what are these costs composed of (costs for data, direct costs for the distribution of data, e.g. for data carriers, costs for the production of photocopies, for staff, a proportion of the costs/flat rate for the generating of data, data documentation costs, etc.)?

ICPSR charges no fees for data transfer to researchers at ICPSR member institutions, and a modest fee (usually in the vicinity of the cost charged to ICPSR for acquisition of the data) to individuals not at member institutions. No extra charge is imposed for a single copy of a file's technical documentation, but we charge a fee for a second or third set of tech doc. That fee is usually the cost of reproduction (for members) or twice the cost of reproduction (for nonmembers).

8. Financing

How is the archive financed? Especially: who bears the costs that emerge for the acquisition of microdata if the data archive's funds do not cover the acquisition/distribution of that data?

For its data activities, ICPSR has two generic sources of funding. The first is membership fees paid by the 370+ institutions that have chosen to affiliate with ICPSR. A second source is grants and contracts for processing and distributing specific data files. A quick glance at the “Topical Archives” section of the ICPSR website (<http://www.icpsr.umich.edu>) will provide a description of the nature and scope of this type of activity. ICPSR's member dues pay for most data acquisition costs (exclusive of those costs covered by the Topical Archives), and nearly all the costs of distribution and long-term preservation. It is our experience that the core work that archives everywhere do (validating statistical data received, providing reliable access to data along with user support, and preserving data on into the future) are under-appreciated and badly underfunded virtually everywhere. In ICPSR's case, universities providing the member dues are

one of the few loci of understanding the importance of these archival operations. Without their financial and intellectual support for this work, ICPSR would have a hard time finding necessary funds to pay for these activities, and would probably have to change its access policies.

9. Demand for official microdata

How strong is the demand for official microdata, how extensive is the provision of consultative services?

There is high demand for official microdata in the U.S., particularly from the scientific community. (Indeed, many of us attribute access to microdata as the key ingredient in the post-WWII revolution in the social and behavioral sciences, and can't imagine what the shape of the social sciences today would be *without* access to microdata.) The two keys to effectively using microdata on a large scale are adequate technical documentation, and provision of consultative services. These two precepts are still honored more "in the breach" (i.e., acknowledged but not supported) than in reality. These lacunae are what create a demand for services like those provided by ICPSR and its sister archives throughout the world.

10. Experiences of the scientific community

Which are the experiences made by academic research concerning the cooperation with the Statistical Office? Where could improvements be made from the scientific community's point of view?

The academic research community in the U.S. finds the statistical agencies to be cooperative and generally sympathetic to research goals. Many of the staff of U.S. statistical agencies are recruited from the academy, and a number of them return to academic ventures after careers in the statistical agencies. The missions of statistical agencies and academic researchers are different enough, though, that some tensions will always exist. The statistical agencies in the U.S. could improve their technical documentation (including providing it in electronic form) in significant ways; that would greatly assist the scientific community's use of official microdata files.