Statistiska centralbyrån Statistics Sweden

Access to microdata in the Nordic countries

Statistics Denmark
Statistics Finland
Statistics Iceland
Statistics Norway
Statistics Sweden

SCB-Tryck, Örebro 2003:02

URN:NBN:SE:SCB-2003-X95BR0301_pdf

Contents

Summary

The protection measures applied to confidential data obtained for statistical purposes are based on several legal acts and directives. The legislation in the Nordic countries is quite similar, with all countries having Statistics Acts regulating the use of statistical information. In principle, information collected for statistical purposes may only be used for the production of statistics and research purposes. However, no specific legal acts deal with Statistics Iceland in cases of access to micro data. Instead, official statistics in Iceland take account of the general acts on data protection, and more recently on acts for the protection of individual's confidentiality.

The National Data Protection Acts also apply to the production of statistics and the release of micro data. The Acts contain rules about the fundamental requirements concerning the processing of personal data.

Data, even anonymous data, obtained for statistical purposes are confidential. In principle, confidential data may be released to a third party only for the purpose of statistical surveys and scientific research. If micro data are released to researchers, the National Statistics Institutes (NSIs) may impose a restriction limiting the researcher's right to use the information. In Sweden however, it is not possible to impose restrictions when data are released to another authority.

Even if legislation in the Nordic countries is similar, there is no common method of releasing micro data from the Nordic NSIs to the users. In Norway, Finland and Sweden, data sets are delivered to researchers that are working outside the statistical offices. Statistics Denmark has another practice. Since their overriding principle is to not release data outside Statistics Denmark, they have set up a scheme with on-site arrangements for external researchers at Statistics Denmark. Statistics Denmark also provides the user access to micro data placed on a special Unix computer in Statistics Denmark, and the user is able to manage this computer from his own office via an encrypted Internet communication.

In the Nordic countries, the same regulations concerning data confidentiality and release of data outside the statistics agency, are also valid when data is delivered to other countries. In practice The Nordic NSIs are restrictive with release of micro data to researchers in other countries. Denmark and Norway does not release micro data to researchers in other countries. Iceland has no experience in delivering micro data to researchers abroad.

Delivery of micro data is usually on an individual (tailor-made) basis from the statistical institute to a researcher, i.e., micro data delivery is not covered by grants provided in the state budget for the direct production of statistics. Consequently, the costs involved in supplying micro data are to be paid by the customers in all Nordic countries. Full costs should be reflected in providing the data, i.e. covering not only direct labour costs, but also rental of the premises, administration, marketing, development and a portion of the joint costs of the statistical institutions for management and administration.

Metadata is an indispensable complement to statistical information today. It is very important that micro data is delivered with satisfactory metadata. The user of micro data must know the content of data, how data is captured, etc. Otherwise there is an overwhelming risk for misinterpretations of statistical results. The central statistical offices are all very much aware of this and are running projects with the purpose of improving metadata.

Every year the Nordic statistical offices compile data sets that are only used for research and analysis activities. In Denmark those data sets produced by Statistics Denmark for researchers shall be archived together with metadata, and after a two-year period other researchers may gain access to these data sets via research facilities established by Statistics Denmark. In the other Nordic countries there is no general archiving regulation corresponding to the one in Denmark. In Norway, Finland and Sweden, however, it is common that special databases with micro data are made accessible for other researchers than those who initiated the database.

The NSIs in the Nordic countries have a number of more or less institutional possibilities of cooperation with their users through "advising boards" for the different areas of statistics.” Access to micro data" is not a field with many formal connections to the users.

During recent years, the Nordic countries, except Iceland, have not experienced any significant mass media debates regarding confidentiality and integrity. In the late 1990s there was an intensive debate on the “Icelandic health-sector database“ that predominated all other subjects that had relevance for the confidentiality issue. However, we do not have much information about public opinion on increased record linkage, development and dissemination of statistical databases.

Other public or private data sources that the work group examined are practising access for other users to micro data very much in the same way as Statistics Norway, Statistics Finland and Statistics Sweden. The study also includes a description of rules and practice for release of micro data in United Kingdom, the Netherlands, Portugal, Canada and EU.

Introduction

A Nordic work group has been set up to describe the Nordic countries' practices concerning confidentiality problems in providing access to statistical micro data for the purpose of research and public planning[1]. The members of the work group have been Birgitta Pettersson, Statistics Sweden, Otto Andersen, Statistics Denmark, Mika Maliranta, Statistics Finland, Magnus Magnusson, Statistics Iceland and Johan-Kristian Tønder, Statistics Norway. The work group has met three times, in Stockholm, Copenhagen and Reykjavik. The purpose of the study is to present a background documentation that is beneficial for national strategies.

National Statistics institutions, NSIs, are required to consider two important issues in their work. It is essential that they:

- Utilize data collection in the best possible manner in relation to the needs of the users. This involves making micro data available to the users.

- Insure protection for individual integrity and confidentiality.

NSIs are dependent on the confidence of the respondents, and they are required to respect confidentiality and protection of individual integrity. This can be problematic with an increased release of micro data that can result in a conflict of interests. While such conflicts may be due to national conditions, there are still many methodical and principle problems common to the Nordic countries. Also, national user solutions for researchers affect the users in the other Nordic countries. Different solutions can result in complaints from the researchers about national NSI practice.

According to the terms of reference, the study should have the following contents:

– The investigation shall create a Nordic description of how micro data is made available for users.

– The main principles of legislation regulating the release of data shall be described.

– The user groups with access to the micro data shall be identified.

– Methods for guaranteeing confidentiality (actual anonymity) shall be explained.

– Pricing principles shall be described.

– In addition to the role of the statistics office, the examination shall also inspect how micro data is made available for other users from other public or private data sources.

[1] To be precise, access to micro data can be provided only for research purposes in Finland. Sometimes research with micro data is done **for** public planning – the analysis of the effectiveness of policy actions, for example.

– The investigation shall give a general description of the attitudes of the public and the media regarding integrity questions, data and statistics. These attitudes should help us to balance the user's needs and research benefits against requirements for confidentiality and integrity. Are confidentiality requirements absolute, or is it possible to evaluate the harm done to the respondents due to the release of non-anonymous data?

– The investigation shall study the cooperation between statistics offices and users and their organisations.

– Discussions should be held about how to assure that metadata and documentation follow research publication, and how data shall be filed.

– An overview shall be made about rules and practice for release to other countries and Eurostat.

The work group was expected to study the problems arising when combining data regarding employer/employee. The group has, however, not been able to report on these very difficult problems within the time limits for producing the report.

The study also includes a description of rules and practice for release of micro data in United Kingdom, the Netherlands, Portugal, Canada and EU.

The System of Official Statistics

The bulk of statistical information originates from various administrative registers. The Population register with the unique identification codes constitutes the key of the official statistical system . Other examples of important administrative registers are Employment registers, Taxation registers (individuals and companies), Job-seeking register and Pension registers. Complementary statistical information is obtained from various surveys. New registers such as the Business Register are constructed and contain data from various registers, providing a framework for various surveys by which more detailed supplementary information is inquired. Furthermore, background information that is contained in register data is being used in the compilation of various statistics.

The unique identification codes for persons, enterprises and establishments, and housing (addresses) used in different registers form the backbone of the NSIs statistical system, whereby different sources of information can be integrated conveniently for various statistical purposes. A valuable feature of the identification coding system is that persons can be linked to companies and to their establishments. The NSIs have an important role in maintaining and co-ordinating this system.

In Denmark, Finland, Norway and Iceland the responsibility for official statistics is mainly centralised. In Sweden, the responsibility for some of the official statistics was decentralised in 1994 from Statistics Sweden to a number of other public authorities. Statistics Sweden is the central administrative authority for official statistics and for other national statistics. Besides Statistics Sweden, there are 24 other public authorities responsible for official statistics. For example, the National Agency for Education is responsible for statistics on education and The National Board of Health and Welfare is responsible for statistics on health and diseases. Statistics Sweden is responsible for more than 50 % of the official statistics. However, Statistics Sweden produce around 80 % of the statistics.

Legislation

Data confidentiality is guided by two major aspects which both are necessary requirements in order to meet the requests from researchers:
(1) general rules (guidelines, screening procedures, contracts, regulations and laws, etc.), and
(2) technical and practical measures for the same purpose.

The legislation concerning confidentiality and protection of individual's integrity is of importance for the possibility for the NSI to provide access to micro-data. The legislation provides the limits for release of data for e.g. research purposes and underpins and constitutes administrative and technical safeguards for legal founding. Specific legislation of importance is the Statistics Act and the Data Protection Acts. To this specific legislation, the current EU legislation with respect to statistical confidentiality should also be added.

EU legislation

The Council regulation (EC) No 322/97 of 17 February 1997 on Community Statistics contains rules that are important for the use of information collected for community statistics. According to the regulation data used by the national authorities and the Community authority for the production of Community statistics shall be considered confidential when they allow statistical units to be identified, either directly or indirectly, thereby disclosing individual information. To determine whether a statistical unit is identifiable, account shall be taken of all the means that might reasonably be used by a third party to identify the said statistical unit. Confidential data obtained exclusively for the production of Community statistics shall be used exclusively for statistical purposes unless the respondents have unambiguously given their consent to the use for any other purposes. However it is possible to allow access for scientific purposes to confidential data obtained for Community statistics.

Of importance for the processing, including release of data, is also the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (the Data Protection Directive). The object of the Directive is to strengthen data protection, e.g. the legal protection of individuals with regard to automatic processing of personal information relating to them. The Directive has been implemented in all the Nordic countries.

The Directive applies to computerised personal data and personal data held in structured manual files. It applies to anything at all done to personal data processing. The new term, “processing”, covers all types of processing of personal data, including registration, storing, disclosure, merging, changes, deletion, etc.

According to the Directive data must be:
– Processed fairly and lawfully.
– Collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. However, further processing of data for historical, statistical or scientific purposes is not considered as incompatible.
– Adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed.
– Accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, with regard to the purposes for which they were collected or for which they are further processed, are erased or rectified.
– Kept in a form, which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed. Personal data can be stored for longer periods for historical, statistical or scientific use.

The legislation in the Nordic countries

The protection measures applied to confidential data obtained for statistical purposes are based on several legal acts and directives. However, it should be noted that access to statistical micro data for research or other purposes is a part of NSIs duty service and is not an obligation given by the law.

In the Nordic countries there are specific and modern Statistics Acts regulating the use of statistical information. Icelandic law with regard to statistical information have not been updated since 1913. However, no specific legal acts deal with Statistics Iceland in cases of access to micro data. In that respect official statistics in Iceland take account of the general acts on data protection and more recently on acts for the protection of individual's confidentiality. In general, given the lack of rules concerning access to micro data, Statistics Iceland is preparing guidelines in order to meet requests for micro data by external users. Statistics Iceland takes notice of two main aspects in cases of micro data: (1) Common rules internationally on good practices for handling of micro data, and (2) specific rules and practices of Statistics Iceland thus far.

Data collected for statistical purposes, in accordance with any prescribed obligation to provide information, or which is given voluntarily, may in principle only be used for the production of statistics. There are exceptions that enable access to data for research purposes and public planning. However, a condition for the use for research is that there is no incompatibility between the purpose of such processing and the purpose for which the data was collected. The processing of data, which includes release of data, must also be in accordance with the regulation concerning protection of individual's integrity.

Besides the Statistics Acts there are specific Personal Data Acts[2] that apply to the production of statistics and the release of micro data. The Acts are based on the Data Protection Directive and contain rules about the fundamental requirements concerning the processing of personal data. These demands include, inter alia, that personal data may only be processed for specific, explicitly stated and justified purposes.

Very stringent rules apply to the processing of sensitive personal data. Sensitive personal data may be processed for research and statistics purposes, provided the processing is necessary and provided the public interest in the project manifestly exceeds the risks of improper violation of personal integrity. Furthermore in Denmark, Norway and Sweden processing of sensitive data for research purposes needs approval. A scientific project involving processing of sensitive personal data is in these countries subject to notification to and approval by the Data Inspection Agency before such processing can commence. This applies to all surveys,

[2] In Denmark - the Act on Processing of Personal Data, Act No. 429 of 31 May 2000 , Finland - The Personal Data Act 523/1999, Iceland - Act on the Protection of Individuals with regard to the Processing of Personal Data 77/2000, Norway –the Personal Data Act (2000), Sweden - the Personal Data Act (1998:204).

whether conducted by a public administration, individuals or enterprises. (In Sweden the approval of the National Data Inspection Agency is not necessary if a research committee has approved the processing.) If the Data Inspection Agency approves the processing, personal data may be provided to be used in research projects unless otherwise provided by the rules on confidentiality. This means that the NSI may take other issues into consideration even if the Data Inspection Agency (or research committee) has approved the processing of data.

Data obtained for statistical purpose are declared as confidential, when they allow statistical units to be identified, directly or indirectly and thereby disclosing individual data. Also anonymous data can be confidential. Statistical data are confidential irrespective of source. Also, data taken from public administrative sources are confidential while in the possession of the NSI. The confidentiality rules are the same irrespective of whether data concerns individuals or enterprises.

Under the main rules, access may be granted in forms which do not allow direct or indirect identification of people or other data subjects such as enterprises. However, confidential data may be released to a third party for the purpose of statistical surveys and scientific research. In Finland it is not generally possible to provide access to data when units can be disclosed directly or indirectly. According to the legislation in Denmark, Iceland, Norway and Sweden, statistical data may even be released with identification data for these purposes. In Finland personal data on a person's age, sex, occupation and education may in exceptional cases be released with identification data for research purposes. One condition in all countries is that access to confidential data for statistical or research purposes must not cause any damage or be detrimental to the data subjects. In practice this means that the NSIs only provide access to anonymous data or de-identified data.

The Nordic countries also have special public business registers that contain some common primary information about enterprises. These registers are (except in Denmark) administered by the NSIs and can also be used for other purposes than statistics or research.

When data has been collected in a voluntary survey the respondents in the statistical surveys must give consent to the release of the data.

It is the NSI that decides whether data may be released for research purposes. However in Norway access for other purposes than statistical must be approved by the Data Inspection Agency. The Agency has given general permission to Statistics Norway to provide access to micro data for research purposes and for public planning. The Data Inspection Agency may nevertheless make exceptions to such obligation of confidentiality for certain types of information if they find it in conflict with the Data Protection Act.

The obligation of confidentiality will also – according to the law or by imposition of a duty of non-disclosure – apply to the recipient of the data. The NSI may also impose a restriction limiting the researchers right to re-communicate or use the information. Breach of confidentiality restrictions is punishable by simple detention or imprisonment. In Sweden, however, it is not possible to impose

restrictions when data are released to another authority. It is therefore important for Statistics Sweden to take into consideration if the data will be confidential according to the Secrecy Act also at the authority receiving data. If not, any one who so desires can have access to the data because of the authority's obligation under Chapter 2 of the Freedom of the Press Act to provide personal data that are not confidential. However, there are rules providing that confidentiality accompanies data to another authority in special situations e.g. if an authority, for research purpose, receives information from another authority where the data is confidential, the confidentiality will apply also within the receiving authority. However, there are no such rules concerning release of data for statistical purposes or public planning.

In Finland a new Act and Decree on the Openness of the Government Activities came into force in 1999. This legislation contains comprehensive provisions on good practice on information management. For instance the Decree includes a detailed list of general data protection measures for confidential data. Statistics Finland like all the other government authorities has to implement these measures by the end of the year 2004.

Providing access to micro data

The general framework in the Nordic countries

In addition to laws and regulations on data confidentiality, all the Nordic countries follow some kind of screening procedure requiring written confirmation that the researcher has signed a general confidentiality statement. Legal contracts are made that include various limitations to the access to micro data by specifying the people, projects, variables and periods during which data can be used in the research. However, as mentioned above, Sweden does not impose restrictions when data are released to another authority.

The Nordic NSIs mainly provide access to micro data to public authorities and people or organisations performing scientific research (universities and research institutions). Statistics Sweden also provides access to micro data to other authorities and municipalities producing statistics. Governmental or municipal institution in Norway can have access to micro data for planning purposes. Generally the uses of micro data for commercial purposes are ruled out.

Methods of providing access

Even if the legislation in the Nordic countries is quite similar, there is no common way of releasing micro data from the Nordic NSIs to the users. How access to confidential data is provided in practice can be divided into a few categories: Off-site access, on-site access, off-line access, on-line access and direct visits (on the payroll) to the keeper of the confidential data.

Off-site access (external use of micro data).

The delivery of anonymous or de-identified micro data outside NSI is the main method by which the NSIs give access to micro data for research in Iceland, Norway and Sweden. Regarding personal data on individuals, the release of anonymous micro data is also the main method in Finland, but micro data on enterprises are released only in exceptional cases. Denmark does not release micro data outside the premises of Statistics Denmark.

Both the anonymous and the de-identified data are in principle available for the researcher for a specified period, for a specified project and for specified people in an institution. When micro data is released outside the NSI it is not possible for the NSI to control the use of the data. The NSI themselves do not inspect that there is no copy available at the institution at the end of the period, or that the data are not used for other purposes than the specified one. The national Data Inspection Agency may observe illegal use of NSI-data on their inspections at the institution, but mainly the framework depends on our confidence in the research institutions and the researchers.

Micro-data is delivered (at least in Sweden, Finland and Norway) to the user by sending out a micro disc by post. The data and the metadata are sent in separate letters.

When micro data is released outside the NSIs, as in Norway, Sweden and Finland (such as samples of individuals), it is common that a pseudo-identifier replaces the identification number. If the user needs annual series of micro data information for the same individuals, the pseudo-identifier may be connected with the identification number, and the combination is to be stored by the NSI. The possibility to have new information added by storing the combination of pseudo-identifier and identification number is restricted to research projects. In Sweden it is also possible to use this method when data is released for statistical purposes. The researcher to whom the data are released must verify that they have a particular need to supplement the information.

In some cases a data set can be delivered outside even when units can be identified directly. This is the case when (a) each and every unit have given the authorisation of the use of data for a specified purpose; or (b) according to the law, e.g. in Sweden and Norway or in Finland where information on individual's age, gender, education and occupation can be delivered outside for the research purposes with identification information. However, one condition is that access must not cause any damage or be detrimental to the data subject. In practice this means that the Nordic NSIs almost never release data with identification information.

Identity data of enterprises including name, postal address and economic sector can be delivered for external use, of course, in the absence of other economic information. This is the case in Iceland, Finland and Norway. In Sweden there is a special register with public data of enterprises. Statistics Denmark is from 2002 no longer responsible for the Danish Central Bussiness Register which is now administred by a special unit in the Central Customs and Tax Administration.

Anonymous micro data, where indirect disclosure is impossible by reasonable efforts, may be delivered outside the NSI. This is typically the case with samples of personal data. The statistics office needs to make sure that the sample does not contain exceptional observations. There must be many observations in each group that can be determined on the basis of the data information. In Finland some samples of company data can be delivered to external researchers only after substantial data protection steps. Due to heavy data protection the usefulness of these data is limited but may, however, serve as a tool for preliminary analysis.

The avoidance of indirect identification is difficult especially when the data set includes a lot of detailed information that may have skewed distribution in statistics, for example. This is the case in the so-called ‘linked employer-employee data’. Statistics Finland has delivered employer-employee data mainly when data contains only very crude categorical information on the employers of the individuals (industry, company size class, etc.).

On-site access (internal use of micro data)

In some cases all-inclusive data protection is very difficult and thus some further limitations may be needed in addition to user screening and legal contracts. In these situations the risk of disclosure is reduced by means of physical limitations. In this alternative a statistical institute makes possible to use micro data by arranging on-site access to external users in a “research laboratory” that is located on the premises of a statistical office. In order to secure data confidentiality identifiers are removed and replaced by pseudo-identifiers. These pseudo-identifiers can be constructed so that a visiting researcher is still able to link separate data sets in a way that is best suitable for the researcher's current purposes.

On-site access arrangement makes it easier to control the appropriate use of micro data. In order to avoid unintended disclosure, external users are usually required to submit all results for inspection before dissemination.

This type of arrangement is was until the establishment of the on-line access the main method used in Denmark for both the individual and business data sets. Statistics Finland has a research laboratory that is currently focused on the various data sets for enterprises that also include quite detailed information on employees. In Norway and Sweden on-site access option is not currently provided. This also applies to Iceland in general except in cases of death certificates, where "bona fide" researchers are allowed direct access to the primary sources.

In Finland some data sets are considered so sensitive and vulnerable to the indirect disclosure that the access is allowed only for employees of Statistics Finland. In these cases it is not a sufficient enough protective measure that the receiver of data has signed a sworn statement. This is the case, when a whole register of individuals is used in the research (employment register data and population census for

example). The same applies for Iceland in cases of population census research. This practice has not been used in Denmark, Norway or Sweden where researchers are granted access to the data needed for a project even for the whole population or all companies.

Off-line access

As an external user normally does not need to see the micro data the programming of the statistical tables and analysis can be made from distance. In this arrangement the user submits (by e-mail) the set-up of his or her computer programs to the data's owner (research laboratory, for example), who then execute the programs and in turn delivers the results to the researcher. This arrangement has earlier been used in Denmark and to a limited extent in Finland but not at all in Iceland, Norway or Sweden. The advantage of this arrangement is for the NSI that it has total control of the use of data as the researcher does not have direct access to them. The disadvantage is for the researcher that the process is more time consuming and probably more expensive than other methods.

On-line access

Since March 1 2001, Statistics Denmark has given most of the Danish researchers the possibility to get access to de-identified micro data on-line.

A research or analysis institute can apply for an authorisation from Statistics Denmark. As of 1 August 2002, 30 institutes had been granted authorisation. Statistics Denmark does not give authorisation to single researchers.

Access is not granted for all datasets; particularly sensitive data (e.g., data on crime) are excluded from the scheme and data on enterprises are assessed carefully to avoid any problems of confidentiality. It is emphasised that the data as main rule consist of samples. If the researchers request access to total populations, the number of variables must be limited. A steering group consisting of the Directors of Statistics Denmark assesses the individual cases. The Board of Governors of Statistics Denmark will assess the scheme regularly.

The technical solution is based on the use of the Internet. The data remain on the Unix servers of Statistics Denmark.

Communication via the Internet is encrypted by means of a so-called RSA SecurID card, a component that secures Internet communication against unauthorised access. In practice the researcher rents a password key (a token) from Statistics Denmark. The token ensures that only the authorised person obtains access to the computer system.

The functionality at external access is largely the same as on the PCs at the on-site arrangement. Data transfer options are not available.

Printouts with (aggregated) tabulations and results from the statistical analysis are sent to the researchers by e-mail semi-automatically (within 5 to 10 minutes). All these e-mails are logged at Statistics Denmark and checked afterwards for confidentiality problems by random sampling by the office of Research and Methods. If problems are found Statistics Denmark makes contact to the researcher in order to avoid dissemination problems and in order to instruct the researcher about the proper way to handle data. At worst Statistics Denmark can cancel the authorization.

The external electronic access by researchers is a valuable step forward for everybody involved. Statistics Denmark remains in control of the data, which are not handed out, and the researchers can work with the majority of the datasets from their own workstations.

Researchers who do not have an authorisation agreement can still use the existing scheme on the on site arrangement for external researchers from Copenhagen or Århus, Denmark.

In Statistics Sweden a system was created in 2001 which make it possible for municipalities to have on-line access to anonymous data.

Principles for pricing

Supplies of micro data are usually considered individual (tailor-made) services from the statistical institute to a researcher (group), i.e. supplies that are not covered by the grants provided in the state budget for the direct production of statistics.

Consequently, the costs involved in supplying micro data are to be paid by the customers in all Nordic countries. The principle is that the full costs should be reflected in providing the data, i.e. cost recovery. The costs involved are to cover not only the direct labour costs, but also rental of the premises, office costs, staff costs, EDP costs, marketing costs, development costs and a proportion of the joint costs of the statistical institutes on management and administration. The Committee has not examined how each individual country distributes its costs, but the necessary overhead contributions are, in practice, added to the actual hourly pay for the staff group in question.

The right to charge service fees differs among the NSIs.

An authority in Sweden is only entitled to charge fees provided that permission has been granted in accordance with a government decision. With respect to Statistics Sweden such permission is given in the Ordinance with instructions for Statistics Sweden. The services are carried out on a compensation basis, and in this light Statistics Sweden has to charge a fee for its services. Although the fees are determined by Statistics Sweden, they have to be fixed in consultation with The Swedish National Financial Management Authority (ESV).

Fees in Denmark are charged according to the rules laid down by the Ministry of Finance Denmark. The rules determine that this activity is to balance financially. If this is not the case, prices are to be adjusted. In accordance with a government recommendation concerning digital management, Statistics Denmark has divided its registers into 3 different types:

1. Registers maintained by an authority that are exclusively used for the production of statistics for which grants are provided in the state budget

2. Service registers that are exclusively established with respect to services

3. Mixed service registers and registers maintained by an authority

In conjunction with fixing a fee per hour, only an hourly fee, (including general overhead) is charged for the first group of registers, whereas an additional fee for, e.g. updating the registers has to be paid in connection with using the other two types of registers. The additional fee is, among other factors, based on an estimate of the number of customers who are expected to use the data. The additional fee is arrived at by adding the true number of hours used for the service in question.

The fee differs depending on the professional classification of the staff. However, Denmark has also decided to use a unit price.

According to a Ministry of Finance decision, Statistics Finland's performances are priced on the basis of market pricing, with the exception of some free-of-charge services. In practise, market pricing allows for flexibility in the pricing of products and services. The services related to the provision of an access to micro data are customer-funded activities, which means that all costs involved should be financed by paying customers. Costs consist of rental costs of premises and facilities needed for research sites, consultation services concerning the data sets, computation costs, licenses etc. The cost of one hour work is 55.5 - 124.5 € depending the status of the worker in Statistics Finland (from keyboader to senior statisitician).In Sweden and Norway the hourly fee is SEK 750-1,000 for the most qualified staff group.

In Denmark and Finland a fee is charged for the use of an on-site scheme as rent of research site.

Metadata

Metadata is an indispensable complement to statistical information today. A difference between “explanatory notes” and metadata should be noted. By metadata we refer to any kind of standardised meta-information.

It is very important that micro data is delivered with satisfactory metadata. The user of micro data must know the content of data, how data is captured, etc. Otherwise there is an overwhelming risk for misinterpretations of statistical results.

The central statistical offices are all very much aware of this and are running projects with the purpose of improving metadata.

In 1994, Statistics Sweden decided that all registers and production systems that Statistics Sweden is responsible for should be documented in SCBDOK. Within Statistics Sweden there is a project concerning development of metadata and documentation, SCBDOK/Metadok. However, the project is not finished and the system is not yet ready.

The main sources of metadata are SCBDOK documentations. SCBDOK contains of a description of the contents of a statistical survey, an observation database and a data processing system. SCBDOK is complemented by product overviews (Metadok). Metadok is a description of variables in an observation database. Furthermore there are quality declarations and other types of documentation available for statistical products produced within the Swedish statistical system. The bulk of metadata will be textual data with limited structuring. These data are most likely to be handled as a text database by free text searchers and document handling systems. Some metadata is available on the Internet

Metadata that has been released to researchers when supplying material (micro data) is quite varied, depending on what the researcher and administrator has agreed to be needed in order to use the material.

Statistics Sweden has also issued a brochure for researchers consisting of a short compilation of the contents of some of the various registers. Another example is the documentation found on the database LOUISE in the form of “background facts for labour market statistics”.

Statistics Iceland is now preparing an Internet facility in order to inform users about the statistical information available. When micro data is considered, no plans concerning metadata as supportive information are scheduled. But if research publication is presented, that publication should be based on the same sources as the statistical field within official statistics. In that case Iceland follows a plan of making “commodity declarations” (by statistical subject) accessible on the Web as a public service which applies to any external scientific research in the same field.

The Danish metadata project TIMES has a goal to make all metadata in Denmark’s Statistics available in a common system by the end of 2002. In a large and ambitious register project, the intention is to reduce the considerable amount of redundant information in the register system, where often the same variable has different designations (names) in different registers. The TIMES system will be available on the Internet by the end of 2002. In practice, however, metadata for many projects will still be supplied on ad hoc basis, such as photocopies of data descriptions. Statistics Denmark has also published the so-called “Declarations of Content” on its Internet site with a description and documentation of Statistics Denmark’s surveys.

Statistics Norway collects metadata from a (somewhat technical) data archive, DATADOK, used in conjunction with the framing of the SAS file. These metadata

are sent to users at the same time the data file is sent. The metadata contains abbreviated names and more complete names, as well as code lists when needed.

Work is now underway for a more “user-friendly” and easily read metadatabase, VARDOK, and testing is planned in conjunction with the Population and Housing Census 2001. Should test results be positive, the new system will be extended to include statistics on individuals and perhaps even economic statistics.

Under each statistics on Statistics Norway*s web-site, they have a general presentation of the variables, the production process and the quality of the statistics. This presentation is available both in Norwegian and English under the heading "About the satatistics".

A lot of metadata are compiled into various function-specific systems in Finland. The overall number of systems is from five to ten depending on the way of defining a system. A problem is that these separate systems are mutually integrated only partially. Currently there are three systems that are mutually integrated; the system of the classification database, the system of the concept database and the Unified file system in order to put electronic data into file archives. The idea is to integrate all the applications where metadata on statistical information, processes, etc., are maintained, e.g. a register of statistical systems and applications. This will be a challenge for an ongoing undertaking that aims to reform and remodel Statistics Finland’s production architecture. An expected result will be an integrated metadata system to which metadata will be recorded once and then disseminated to particular applications that produce information services. A lot of metadata about statistics have been made available to users on the Internet.

When it comes to micro data sets developed for research purposes, different types of information about data is compiled and documented in various ways. The user-friendliness, quality and quantity of metadata that is made available to researchers will improve when the integration of various systems is carried out.

Reuse of research data

Every year the Nordic statistics offices compile various (numerous) data sets that are only used for research and analysis activities. These data are sometime paid for by public funding, e.g. in Denmark. These data sets could be used to a much greater extent if other researchers were given access to them. It should also always be possible for other researchers to review the produced research results without encountering an economic burden.

This is why there is much to say for availability of costly data sets of a certain size to researchers for reuse.

In Denmark it is customary with this kind of general reuse. Those data sets produced by Statistics Denmark for researchers shall be archived together with metadata, and (as the general rule) after a two-year period other researchers may

gain access to these data sets via research facilities established by Statistics Denmark.

The actual establishment and operation of this reuse system is expected to be financed by compensation to Statistics Denmark from the Ministry of Research.

Danish research groups are in considerable disagreement about the time period before other researchers can gain access to the data. Some researchers feel one year is sufficient while others think there should be at least a three-year requirement so as to protect individual researchers against other researchers' use of the data.

In the other Nordic countries there is no general archiving regulation corresponding to the one in Denmark.

In Norway, however, it is common that special databases with micro data financed by research councils or ministries are made accessible for other researchers than those who initiated the database. This applies to cases in Norway where SSB files (stores) data. Even in Sweden there are special databases set up to fill the needs of researchers. Individual researchers can gain access to data from these databases on those terms that apply to release to separate projects. There may also be an update of the released material.

Data sets specially produced for separate projects are not normally filed (stored) by Statistics Sweden. In some cases these data are filed (stored) at the research institution that received the data set.

Backup copies from main files are recorded frequently. A researcher is allowed to hire disk space for recording those personal data set files (for example in SAS format) in the computer or CD that she or he may need during or after the project. CDs are kept in a safe box.

Cooperation between the NSIs, the users and their organisations

The NSIs in the Nordic countries have a number of more or less institutional possibilities of cooperation with their users through "advisory boards" for the different areas of statistics. "Access to micro data" is not a field with many formal connections to the users.

Statistics Norway has an agreement with the Norwegian Social Science Data Services (NSD). NSD is a part of the Norwegian Research Council. Researchers that ask for micro data need information on how to get such data and the rules and procedures for getting them. NSD can give researchers information and forward their demands to Statistics Norway. If the user is satisfied with anonymous data, Statistics Norway can ask NSD to make anonymous data for the researcher out of de-identified data Statistics Norway are giving NSD access to for a period. If the user wants access to de-identified data, NSD can help him to get admission from the Data Inspection Agency to handle such data. The administration and practical

work of giving access to de-identified micro data has to be done by Statistics Norway.

Statistics Iceland has no special arrangement for cooperation with the users concerning micro data.

Statistics Denmark collaborates quite closely with an institution: National Centre for Register-based Research which has been established and funded by a research foundation: The Danish National Research Foundation. The aim of the centre is to promote and improve the use of Danish registers within the research areas of both social sciences and health sciences. The centre is located in Århus, Denmark, in collaboration with a small department of Statistics Denmark.

In August, 2001 a work group published a Report: "Registerforskning – Enestående danske muligheder" with an analysis of the barriers to Danish register research. As a consequence of this analysis it is expected that Statistics Denmark from 2002 will have some funds (6 mill. d.kr.) allocated in order to make some data cheaper, but especially to pay the cost of running the on-site and on-line arrangement.

The NSIs in Denmark, Finland and Sweden have an advisory scientific board with the aim to give advice to the NSIs in different questions. For example, the board advises about methods and quality in the production and the policy of using micro data.

Public opinion

The confidence of the public is of major importance for the NSIs. The Nordic NSIs are therefore naturally concerned about how the increasing demands for micro data for research purposes will influence the public opinion. However, no serious confidentiality accidents concerning data released to researchers have occurred in the Nordic countries.

During recent years, the Nordic countries (Iceland excepted) have not experienced any significant mass media debates regarding confidentiality and integrity. However, this does not mean that the public accepts an increased use and dissemination of statistical information. Actually, we do not have much information about public opinion on increased record linkage, development and dissemination of statistical databases. A study conducted by the Norwegian Data Inspection shows that 70% of the Norwegian population feel that integrity issues are significant and that it is important to obtain knowledge about which information is available on oneself.

The intensity of the debate in the late 1990s about the "Icelandic health-sector database" predominated all other subjects that had relevance for the confidentiality issue. The confidentiality questions which are relevant for Statistics Iceland are the outcome of the twofold character of the office, as an administrative register (the National Population Register and the Register of Enterprises) and secondly as an

office of public statistics. Of these two sides the administrative one is more vulnerable and debated than the statistical counterpart. In short, statistics with regard to confidentiality are not of primary concern in the mass media, mainly because of the restricted access of micro data granted to external users.

Recently in Sweden, however, there has been some debate about integrity issues in connection with a proposal on a new population and housing census. Critical viewpoints have been presented over the years regarding population and housing censuses in the past, as well as the register-based population and housing census now. One of the questions has been the risk for invasion of privacy. Concern has been expressed about the creation of new large registers with information on the total population.

The Norwegian media has posed questions about personal integrity in connection with commercialisation of personal information, surveillance and control of individuals, as well as focusing on people in marketing, research and the media. When it comes to the position on personal information for statistical purposes there is a certain duality. On the one hand, the media is interested in the statistics, while on the other hand they are also interested to find negative angles regarding the dangers of storing data for such purposes. For the most part, however, the media regards the collection and storage of data in Statistics Norway in a neutral manner.

Every other year, Statistics Sweden conducts a survey to find out the position of the public regarding Statistics Sweden and statistics. Of those who answered the survey in 2000, 80% felt that statistics were important. Only 9% felt there was a risk that information submitted to Statistics Sweden could be exposed in an unfavourable way. The public appears to have confidence in Statistics Sweden and our handling of confidential information. Users of material from Statistics Sweden such as researchers, analysts, and statisticians rely on Statistics Sweden to maintain respondents' confidence. One study from Norway showed that nearly 70% of the Norwegian population have great confidence in Statistics Norway. However, the survey does not reveal whether this confidence is due to the statistical results and research, or due to work for protection of personal privacy. The public and the press in Finland have a rather positive attitude to official statistics. In Finland a survey showed that 84 % of the population have confidence in Statistics Finland and 70 % think that official statistics are impartial. Surveys concerning the attitude to e.g. authorities indicate that the press in Finland thinks that Statistics Finland is one of the most reliable organisation in the public civil service. Statistics is an important source for articles in the press.

However, this faith in the NSIs has not led to massive support for statistical surveys. For example, in recent years Norway, Iceland and Sweden have experienced a lower response rate for labour force surveys. However, this lower response rate is probably due to factors other than integrity issues.

Providing access to micro data to researchers abroad

In the Nordic countries the same regulation concerning data confidentiality, as for release of data outside the NSIs, are in principle also valid when data is delivered

to other countries. There are however some restrictions. According to the Data Protection Directive it is in principle forbidden to transfer personal data that is being processed to a third country (a country outside the EU and EEA) unless the third country in question ensures an adequate level of protection. The Data Protection Acts in the Nordic countries contain similar rules about release of data to a third country.

In Sweden the Secrecy Act is also of relevance. According to Chapter 1 section 3, the release of confidential data to an authority or an international organisation outside Sweden is not allowed unless it is communicated in accordance with special provisions in legislation. Also, the information in a corresponding case might be given to a Swedish authority and the authority holding the information deems it evidently compatible with Swedish interest that the information is communicated. The EU regulation is such special provisions that make it possible to release micro data to Eurostat. There are no other special provisions concerning statistical micro data.

In Sweden the release of micro data to an authority in other countries for research is therefore possible only if it is compatible with Swedish interest that information is communicated. Micro data may be released to private researchers in other countries if it is evident that the information can be disclosed without the person whom the information concerns suffering loss or being otherwise harmed. In practice Statistics Sweden is restrictive with release of de-identified micro data to researchers in other countries.

Regards to the Statistical Act in Norway, all users of microdata are bound to secrecy. Since the legislation is not valid outside Norway, and Statistics Norway is thus not able to control if researcher in other countries maintains the confidentiality rules, Statistics Norway find it indefensible to release micro data outside Norway. However, the legislation accept transfer abroad if Norway is subject to an obligation to make a transfer pursuant to an international agreement or as a result of membership of an international organization.

In Finland the same regulations concerning data confidentiality as in Sweden, as for release of data outside Statistics Finland, are valid. An applicant must provide a description about how the data confidentiality is secured in the recipient country.

Denmark does not release micro data to researchers in other countries but foreign researcher can use the Danish on-site arrangement under the same conditions as Danish researchers. Iceland has no experience in delivering micro data to researchers abroad.

Eurostat

The release of information to Eurostat is regulated in the EU regulations on statistics. According to Regulation 1588/90[3] the national authorities shall be authorized to transmit confidential statistical data to Eurostat. National rules on

[3] Council Regulation (Euratom, EEC) No 1588/90 of 11 June 1990 on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities

statistical confidentiality may not be invoked to prevent the transmission of confidential statistical data to Eurostat where an act of Community law governing a Community statistic provides for the transmission of such data. This means that NSIs in principle are bound in regulations to release micro data for community statistics. However, transmission of data which are not covered by a specific Community legislative act is voluntary and that national rules can prevent the transmission of confidential data. Transmission of confidential statistical data shall be carried out to Eurostat in such a way that statistical units cannot be directly identified. This does not preclude the admissibility of more far-reaching transmission rules in accordance with the legislation of the Member States.

Policies and practices of other authorities and institutions giving access to micro data

The Nordic NSIs use considerable amount of data from administrative registers at other authorities. These authorities often also have the right to provide access to micro data for research purposes. It has not been possible to examine how each authority provide access to micro data. However, the NSIs normally provide micro data to researchers from different administrative registers. The work group has therefore concentrated on micro data on social security and health. These data are normally released for research purposes from the authorities responsible for such data.

National Insurance Administration (Norway), National Research and Development Centre for Welfare and Health (Finland) and the National Board of Health and Welfare (Sweden) have a number of micro data on social security and health. They provide access for other users to micro data very much in the same way as Statistics Norway, Statistics Finland and Statistics Sweden. However the National Board of Health and Welfare in Sweden are more liberal with release of micro data with identification data. For medical research the board provides access to data with identification information if the researcher has a specific need for such data. National Research and Development Centre for Welfare and Health in Finland may also release health data with identification information for medical research. A planned Icelandic health-sector database seems to have more precautionary actions than micro data in Statistics Iceland.

In Denmark the National Board of Health is next to Statistics Denmark the most important institution delivering micro data for research. The data are primarily about cause of death and on hospital admissions and are normally delivered off-site contrary to the practice in Statistics Denmark. Statistics Denmark has access to the data mentioned and is allowed to link the data to its own data for specific research projects.

In Finland some private firms maintain large company data sets that are increasingly used in the scientific research. Comprehensive payroll data on manufacturing workers is another important data set in Finland. It is compiled and maintained by The Confederation of Finnish Industry and Employers. The employers' organisation has delivered this data to Statistics Finland for the purpose of producing official statistics for many years. More recently the organisation has

provided access to this data source for wider user groups in research institutes. These data are particularly valuable in economic research as it can be linked to company data sets sold by private firms.

In Sweden there are some private institutes that collect data from enterprises. The institutes mainly use the data for their own research or for analysis purposes and do not provide access to the data for researchers outside the institute. However, some of these institutes are cooperating with Statistics Sweden and have delivered data for the production of official statistics.

Rules and practices for release of micro data in UK, NL, Portugal, Canada and EU

United Kingdom

There is no single Statistics Act in United Kingdom. Legislation governing the release of statistical micro data is specific either to the Act under which the data was collected, or specific to the type of data being processed. The most important legislation that relates to the type of information being processed is the Data Protection Act (1998). As in the Nordic countries, the Act contains rules about the fundamental requirements concerning the processing of personal data.

The main principle for the disclosure of micro data concerning businesses is that no individual estimates or returns and no information relating to an individual undertaking shall be disclosed, without the previous consent in writing of the person carrying on the undertaking which is the subject of the estimates or information. Census data, health data and civil registration data are also confidential and as a main principle, it applies that no information with respect to any particular person shall be disclosed.

The National Statistics´ (ONS´) Code of Practice emphasises the commitment of the National Statistician to maintain the confidentiality of individual people, businesses and organisations. The Code requires that "Individual statistical survey or census data will be used for statistical purposes only, unless specified by statute or agreed by respondents”. Their confidentiality will be absolutely respected.

The current practice varies considerably across the different business areas of ONS. There are a very large number of successful relationships between ONS and academic, government and commercial statistical researchers, the arrangements for which have emerged and evolved over time.

Release of, or access to, ONS micro data may be allowed through the Public Records Office, as may be required by legislation, to the academic or commercial community through a partner organisation or through the National Statistics online service 'Statbase'. Release of micro data may also be made to specific users of statistical information, to other central or local government departments for statistical purposes, to Eurostat as required by European law, or through archiving arrangements organised on an ad hoc basis for a specific and limited statistical purpose, subject to specific access arrangements.

Methodology Group in ONS has a cross-cutting role in providing disclosure control, data sharing and confidentiality solutions and support to National Statistics business areas. These methods will become best practice for all Official Statistics. Protection measures are designed according to the content of the micro data and the status of the recipients. Techniques employed include 'anonymisation' through removal of direct identifiers, sub-sampling (release of a random sample of records) record-swapping suppression of categorical variables collapsing, top/bottom-coding, perturbation or rounding of continuous variables.

The ONS Executive considers access to ONS potentially disclosive micro data to be a privilege gained through trust, and not a right. For each release of disclosive micro data, therefore, there must be an access agreement, and an assessment of the ability of the recipient to maintain its conditions. All new or changed micro-data releases must pass through a Micro Data Release Committee which for approval. Proposals reach the Committee as a completed proforma, which will form a repository of approved micro-data release methods.

The Netherlands

In the Netherlands the current Statistics Act get into force in 1996. There are also the Law on Economic Statistics and the law on external use of the General Business Register. There is a draft for a new Dutch Central Bureau of Statistics CBS law that is meant to supersede all earlier statistical legislation. The draft legislation is currently with Parliament.

Information collected by the CBS shall only be used for statistical purposes and no individual data whatsoever may be given to anyone (including other government services) outside the CBS, without explicit written permission from the data provider. The Statistics Law regulates two exceptions to the general confidentiality rule: controlled supply of micro data for research purposes and the supply of confidential data to Eurostat.

It is possible to release micro data under contract to specified users. The contract implies that the data concerned will not be matched to other data, and will not be transmitted to other researchers. Researchers will show the draft of their results (aggregates only) to Statistics Netherlands for a disclosure check, and they will destroy data they do not use any longer.

There are four categories of users, the universities, the Dutch government planning offices, Eurostat and research institutes. The research institutes may be authorised by the Central Commission for Statistics. The Commission has formulated its policy, which states that applicants must have legal personality and continuity, must conduct research as a core activity, and are required to publish.

In 1994 an agreement concerning access to micro data for researchers was concluded between the CBS and The General Board of the Netherlands Organisation for Scientific Research (NWO) about a ‘Scientific Statistical Agency’ (SSA). The SSA deals with micro sets about people and households exclusively.

Data sets about business surveys are excluded. The CBS has decided to create CeReM (Centre for Economic Micro data Research) for a trial period of three years. CeReM is a facility for 'on site' analysis of micro data relating to business surveys. The Central Commission for Statistics defines the general criteria for those categories of researchers allowed to work 'on site'. Users must sign a confidentiality statement, output can only be exported after a 'disclosure protector' of the CBS has checked it and users work on stand-alone machines only, so they cannot e-mail from their work station. CeReM functions satisfactorily and there is a growing interest to work 'on site'.

Currently, CBS release micro data from social sample surveys on CD Rom under contract for statistical and research purposes only, and access on site to business micro data and other (administrative) micro data is given to bona fide researchers, The CBS is going to experiment with remote access and remote execution facilities for both types of data.

Portugal

In Portugal there are several rules concerning micro data. According to the basic law of the National Statistical system all statistical information of an individual nature collected by NSI is confidential in nature. Individualised information on individuals may never be divulged. Information about cooperatives, companies, credit institutions and other economic agent may not be divulged, unless with the written authorisation of the respective representatives or after authorisation has been given by the Higher Statistical Council, case by case, whenever the needs of planning and economic coordination or external economic relations are involved.

In 1999 INE Portugal signed a protocol with the Ministry of Science and Technology to provide researcher with anonymised microdata and metadata for 19 INE projects. The protocol provides for the availability of historical data, which covers the last decade and the last two decades in certain cases. In 2000 INE-PT decided to create a statistical Information system for Researchers with Internet access. The system, which is currently undergoing testing, is available on the private network of the Ministry of Science and Technology. The objective is to provide data to researchers accredited by the Science and Technology Observatory. Before data is made available for external access, it must be validated and documented by the project supervisors. After approval by the project supervisor the data is transferred for external access.

Canada

Public Use Microdata Files (PUMFs)

Revisions to the Canadian Statistics Act in 1971 made it possible for Statistics Canada to produce Public Use Microdata Files (PUMFs). These provide the most cost-effective method of access, have been very popular with the research community and largely account for the success of Canada's Data Liberation Initiative. Through this data consortium, 66 Canadian universities have access to

all of Statistics Canada's available PUMFs for teaching and research. Statistics Canada has created more than 350 PUMFs since the outset of the program.

The survey master files are processed in such a way as to reduce the probability of disseminating information about unique respondents. The survey managers are responsible for carrying out the disclosure control procedures by reducing the geographic detail, by suppressing variables and by collapsing response categories. The resulting public file is reviewed by the Microdata Release Committee consisting of a diverse range of senior managers. The survey managers are also expected to produce a codebook which documents the survey.

It was not until the agency began creating more complex surveys, especially longitudinal surveys, that it was felt that the public files with sufficient detail to be analytically useful, could not be produced.

Research Data Centre

The emergence of complex longitudinal files such Canada's National Population Health Survey (NPHS) has made it virtually impossible to produce a PUMF file. To compensate for this loss, and in an effort to strengthen Canada's social research capacity and to support the policy research community, nine Research Data Centres have been established in Canadian Universities. RDCs operate under the privacy and confidentiality provisions of the Statistics Act and are staffed by a Statistics Canada employee at all times. This also necessitates review and approval procedures that other forms of microdata access do not, and is an extremely expensive solution.

Remote Job Submission

For researchers who cannot travel to the RDCs, remote job submission procedures are one attempt to provide microdata access for researchers. A remote access program's success is contingent on the availability of good survey documentation, the creation of synthetic (dummy) files for program testing, the ability to run a variety of software and a relatively fast turnaround time.

Resident researchers

Accommodating 'resident researchers' with Statistics Canada's regional offices and headquarters has been an informal program that has been used in fairly specific circumstances. The increasing availability of other means of access will likely reduce this an access method.

EU

Eurostat may allow researchers access to confidential data for scientific purposes. In principle, such access may be granted under Article 17 of Council Regulation (EC) 322/97. It is specified, however, that there must be explicit approval for such access from the Member States that supplied the data to Eurostat.

The European Community has recently approved a new regulation concerning access to confidential data for scientific purposes. (Commission Regulation on 17 May 2002. EC No 831/2002 implementing Council Regulation (EC) No 322/97 on Community Statistics). The regulation contains rules concerning conditions for when Eurostat may allow access to confidential data from four specific surveys. There is no demand for an explicit approval from the member state, but the member state must be informed and has the possibility to deny release of data to a specific researcher/research project. The Regulation applies to four specific datasets, The European Community Household Panel, The Labour Force Survey, The Community Innovation Survey and The Continuing Vocational Training Survey.

In the regulation, the screening procedure specifies especially who is provided access to confidential data. According to article 3 access to confidential data may be granted (under given conditions) to researchers of specified bodies.

Secondly, the general conditions are outlined by stressing the importance of making legal contracts and other practical preparations (article 4).:
(a) an appropriate request together with a detailed research proposal in conformity with current scientific standards has been submitted;
(b) the research proposal shall indicate in sufficient detail the set of data to be accessed, the methods of analysing them and an indication of the time needed;
(c) a contract specifying the conditions for access, the obligations of the researchers, the measures for respecting the confidentiality of statistical data and the sanctions in case of breach of these obligations has been signed;
(d) the national authority, which provided the data, has been informed before access is granted.

Thirdly, the supervision procedure, the specific requirements, of the research projects is described (articles 5 and 6):
(a) the research will be carried out exclusively within the premises of the Community authority and under the supervision of a designated official of that authority;
(b) the results of the research shall not leave the premises of the Community authority without prior checking to ensure that they do not include confidential data;
(c) prospective results to be published or otherwise released shall be checked by the Community authority to avoid disclosure of confidential data.

Glossary and definitions of terms

Anonymous micro data: Individual statistical records which have been modified in order to minimise, in accordance with current best practice, the risk of identification of the statistical units to which they relate.
Applicable data protection law: The legislation protecting the fundamental rights and freedoms of individuals and, in particular, their right to privacy with respect to the processing of personal data applicable to a data controller in the NSI in which the data exporter is established.
Confidentiality/secrecy: Prohibition on disclosing information by making an official document available to an unauthorised person.
Consent: The acknowledgement by an individual (generally in writing) that he/she gives permission for participation in a certain study, or that information on him/her may be used, the individual having received and understood information on the purpose and practice of the study, the potential benefits and risks of participation in the study, and that he/she is free to refuse participation or cease participation at any time. Consent is an appeal of a statistical office to take seriously the relationship with "data subjects" while keeping in mind the objective of providing useful and valid statistical information. The relationship with the "data subject" becomes more complicated if the collector of statistical data intends to release micro data and/or intends to combine data from various sources, administrative and others. Definition according to EC Directive 95/46: "Any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed. **Cryptography**: "The science that deals with the design of encryption or cipher systems." Cryptography can help to solve problems related to "statistical disclosure control techniques that are not addressed by "statistical data protection".
Data protection, classification of: A practical basis for the protection of statistical data of a statistical institution. "Its purpose is to classify statistical material into different groups according to its confidentiality (sensitivity). The main classification standard is the sensitivity of the data from the viewpoint of personal integrity, the integrity of business enterprises or a national safety.
De-identified micro data: Micro data where official identification codes, names, addresses, etc. are removed.
Direct identification: Identification of a statistical unit from its name or address, or from an officially allocated and published identification number.
Disclosure avoidance: Abbreviation or distortion of the data "sufficiently to eliminate all relevant risk of identifying individual respondents. The challenge posed by the confidentiality problem is to provide protection to respondents without damaging the completeness, accuracy and usefulness of the data products."
Disclosure of data: Uncovering of data; reveal of data.
File: Any structured collection of micro data where data on certain individuals, establishments or firms can be located.
Identifiable person: A particular individual (or a "data subject") that can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental,

economic, cultural or social identity. Identifiable plant or firm can be characterised in a very much analogous manner.
Identifiers (direct): such as ID-numbers, names and addresses, where recorded.
Indirect identification: Possibility of deducting the identity of a statistical unit other than from "direct identification".
Micro data: Unit record or low level data that has not been sufficiently aggregated to be made publicly available. Synthetic or quasi unit record datasets, which require a statistical office to have regard to similar issues of confidentiality, is also considered within this context.
Off-site access: The researcher has custody of a dataset for analysis.
Off-line access: In this model a researcher email programs to the data host, who then runs them in the protected environment, make sure that units cannot be disclosed indirectly and then send the output to the researcher.
On-line access: Access to micro data held in a safe setting allows academics to analyse registers and survey data sets that are not allowed to be distributed. In this model, researchers have on-line access (from their own workplace) to data through an encrypted Internet connection. The researcher is only allowed to run computer programs resulting in tabulations and analysis. Data transfer is prohibited.
On-site access: The researcher analyses datasets while in a protected environment or a designated space on the premises of a statistical office. The statistical authority describes the conditions of use.
Personal data: Any information relating to an identified or identifiable natural person ("data subject"), i.e. information that can be directly or indirectly traced to a particular individual.
Processing of (personal) data: Any operation or set of operations which is performed upon (personal) data, whether manually or by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction.
Pseudo-identifier: The replacement of real or direct identifiers (e.g. person numbers) with pseudo-identifiers (e.g. information less integers created by a table or by encryption of the real identifiers) in connection with the “anonymization” of micro data, before they are released to external users. A translation key between pseudo-identifiers and real identifiers may be created and kept by the responsible statistical agency.
Public planning: Preparations by an official administration for political decisions in local or national assemblies with a long run effect. The administration may use of research institutes or consultants to make studies ("utredninger") on some aspects of the plan. such studies are to be regarded as a part of "public planning".
Safe custody: A statistical office "must be assured that the data is going into safe custody before access is granted. This is done by either controlling the environment within which the data is accessed or by allowing it to go into the custody of an agency or an individual who will maintain the protection of confidentiality that" the statistical office pledges to its respondents.
Scientific research: "Study of phenomena relating to groups of individuals or enterprises in the general public interest and the results of which are made publicly available." The research process separates the information from the person: personal data are collected and processed with a view to producing consolidated

and anonymous results; therefore, the relevant information cannot, under any circumstances, entail direct or individualised consequences for individuals.

Sensitive data (regarding individuals): "Sensitive data include: (a) Data revealing a person's racial or ethnic origin. (b) Data revealing political opinions, trade union membership, religious beliefs and other convictions. (c) Data revealing on whether a person has been suspected of, charged with, indicted for, or sentenced on account of a punishable offence. (d) Health data, illness or disability. (e) Data on sexual behaviour of a person. (f) In Iceland also Data on social services, economic support, social assistance and related social welfare services received by a person.

Statistical confidentiality: The protection of data related to single statistical units, which are obtained directly for statistical purposes or indirectly from administrative or other sources against any breach of the right to confidentiality. Statistical data shall be considered confidential when they allow statistical units to be identified, either directly or indirectly, thereby disclosing individual information.

Statistical data protection (SDP): A wider concept than "statistical confidentiality" and takes into account all three steps of production. SDP is multidisciplinary and draws on computer science (data security), statistics and operation research.

References

1. The National Statistics Code of Practice and Protocol on Release Practice, (The National Statistics, ONS).

2. Statistics, Secrets and Sediments, by Willem de Vries and Joris Nobel, Statistics Netherlands (Wiesbaden 1999).

3. Data in the Polder, by Joris Nobel, Statistics Netherlands (Luxemburg 2001).

4. Statistical information system for researchers. Paper presented by Fernanda Perpétou, Statistics Portugal, 19th Ceies Seminar, Lisbon, 26 and 27 September 2002.

5. Solutions to provide access to micro data. Paper presented by Ernie Boyko, Statistics Canada and Wendy Watkins, Carleton University. 19th Ceies Seminar, Lisbon, 26 and 27 September 2002

Statistics Sweden
Box 24300
S-104 51 Stockholm
phone: +46 8 506 940 00

Contact:
Birgitta Pettersson
birgitta.pettersson@scb.se
+ 46 8 506 941 32

www.scb.se