

THE DANISH SYSTEM FOR ACCESS TO MICRO DATA

To facilitate register-based research, Statistics Denmark gives researchers access to de-identified micro data, i.e. data on person, family, household, workplace or company level. Access is given as secure remote access to servers at Statistics Denmark from the researcher's own computer through the Internet. This arrangement gives researchers a unique possibility to use micro data in their research.

The present paper briefly presents the rules and the organisational framework for researcher's access to micro data at Statistics Denmark.

According to these rules access to micro data can only be granted to researchers and analysts in Danish research environments after approval from Statistics Denmark. Foreign researchers affiliated to a Danish authorised environment are given access to Danish micro data as well.

1. Short historical background

Denmark introduced the Personal Identification Number (CPR) in 1968 and it was used in a census for the first time at the Population and Housing Census in 1970. Accordingly, this became the first Danish register using the Personal Identification Number as an identification key. During the 1970s the first attempts were made to base the production of statistics on registers. In 1976 a register-based population census was conducted as a pilot project, but the registers were not sufficiently comprehensive and well-established until 1981, when a proper register-based population and housing census was conducted containing most of the conventional population and housing census information.

Like in the other Nordic countries, the person and business registers in Denmark cover today a very substantial part of the production of statistics. Most of the registers date back to the beginning of the 1980's, which make them very relevant for, e.g. longitudinal research projects. The contents of the registers also cover many fields of research such as labour market research, sociology, health and business economics. The strength of the system is that the identification keys (person number, address, central business register number and property title number) enable correlation of data both within a specific year and, as mentioned, longitudinally across several years.

2. Rules for researchers' access to micro data

Access to micro data can be granted to researchers and analysts from research environments pre-approved by Statistics Denmark.

Authorisations can be granted to public research and analysis environments (e.g. in universities, sector research institutes, ministries, etc.) and to research organizations as a part of a non-profit foundation.

Within the private sector, the following user groups can be granted authorisation if they have a stable research or analysis environment (with a responsible manager and with a group of researchers/analysts):

1. Non-governmental organisations
2. Consultancy firms
3. Enterprises. However, single enterprises cannot have access to micro data with enterprise data

In order to grant an authorisation, Statistics Denmark will evaluate the proposed organization carefully and, especially when it is an organization or firm within the private sector, Statistics Denmark takes the credibility of the applicant into consideration (among others, ownership, educational standard among the staff and the research experiences).

Statistics Denmark will not grant authorization to single persons.

Access to data is given according to a so-called “need to know”-principle, meaning that researcher can only get access to the data needed to fulfill their research purpose.

3. Foreign researchers

Only Danish research environments are granted authorisation as Statistics Denmark is not able to enforce a contract effectively abroad. Foreign researchers are, however, given access to micro data through an affiliation to a Danish authorised environment. Affiliation can only take place if the authorised environment is willing to take the responsibility for the foreign researcher making sure that all existing rules governing access to micro data are observed. The research environment must also appoint a contact person who will undertake the responsibility for all contact between the foreign researcher and Statistics Denmark.

If a foreign researcher is interested in getting access to micro data, Statistics Denmark recommends the researcher to establish a contact to a Danish authorised environment and discuss the possibilities to get affiliated.

In Appendix 1 the rules for the researcher’s access to anonymised micro data under Statistics Denmark’s research schemes are laid out.

4. Organisational framework

The scheme is administered centrally by the Division of Research Services at Statistics Denmark. The staff of this division also creates a substantial part of the interdisciplinary datasets and has access to all relevant data at Statistics Denmark in order to reduce the administrative and bureaucratic work. The scheme requires close cooperation between the Division of Research Services and the individual statistics divisions. The advantage of this central organisation is that the individual researcher is fully aware of whom to negotiate with and who is responsible for the dataset supplied.

5. Data available for researchers

Most of the data derive from the administrative registers of governmental agencies, offering high data quality for the entire Danish populations of persons, buildings and companies. Data can be combined in endless ways allowing researchers to produce

unique analysis of dynamic processes and fluctuations, using the Danish population as their study population.

All together, data from 250 subject-areas are available for research purposes through Statistics Denmark. The majority of the registers fall into the following headlines:

- Population and demography
- Labour market
- Earnings
- Income, consumption and prices
- General economic statistics
- Agriculture
- Manufacturing industries
- Construction and housing
- Service sector
- Transport
- Environment and energy
- External trade
- National accounts and balance of payments

All statistics have an affiliated quality declaration, offering a comprehensive description of the sources and methods used. The declarations describe quality and content of data. For each statistic, we provide information about purpose, history, collection method, accuracy, reference period, and comparability over time. For further information please see: www.dst.dk/documentation.

Data from Statistics Denmark can easily be linked to data from other sources, e.g. survey data or data from other governmental agencies such as Danish Health and Medicines Authority, etc. Statistics Denmark also carries out interview surveys customized to users' needs.

In cases where data from other sources than Statistics Denmark are included in the research project, approval by the Danish Data Protection Agency is needed before such processing can commence.

To reduce the cost of datasets for research purposes and solve some specific data problems, Statistics Denmark has set up a number of research databases. These databases are hardly ever used in the actual production of statistics, but are a kind of intermediate products for the benefit of the research process.

The most popular register in health research is **The Register of Medicinal Product Statistics** holding information on doctors' prescriptions of medicine sold by the pharmacies in Denmark. The researcher's access to this register has boosted the interest for register-based epidemiology research. The register contains information on the Danish population's use of medicine from 1995 and onwards.

A frequently applied research database is the **Integrated Database for Labour Market Research (IDA)**. One reason for creating the database was to solve a difficult problem of definition: Identity of enterprises over time, a task that individual researchers were unable to handle for reasons of both time and funding. Nine to ten man-years were spent on the development, which was jointly funded by the Danish Social Science Research Council and Statistics Denmark. Since the establishment of IDA, Statistics Denmark has handled the updating of the database financed by user charges. The IDA-database contains information on the total population of people and enterprises in Denmark from 1980 and onwards.

The **Fertility Database** contains all references between child, mother and father found in the basic population registers from 1973 and onwards. In the register we have complete references between mother and child for children born from 1960 and

onwards. The Fertility Database, which contains information of all people in the fertile age groups, is widely used in health research.

6. Working with micro data on the research servers

Micro data for research purposes using registers from Statistics Denmark must not be handed over to the researcher. All access to micro data for research is, therefore, given through a powerful research server placed at Statistics Denmark. The research server is separated from the production network and only contains de-identified micro data for research purposes.

When research data for a project have been prepared by the Division of Research Services, the data are transferred to the research server where remote access is given via the Internet. The researcher has to sign an agreement with Statistics Denmark before access is granted, cf. the agreement in Appendix 2. Communications via the Internet is protected and encrypted, which secures against unauthorised access.

The agreement, which has to be signed by the researcher before getting access to data, states that all work on micro data must take place on the research server and no attempts to identify people or enterprises – or to remove micro data must be made and is considered a very serious breach of the agreement between the researcher and Statistics Denmark. Only aggregated data, where no identification of persons or enterprises is possible, can be removed from the server.

All aggregated results from the researchers' computer can be stored in a special file and such printouts are sent to the researchers by e-mail. This is a continuous process (every five minutes) and has shown to be quite effective. The advantage for Statistics Denmark is that all e-mails are logged at Statistics Denmark and checked by the Research Service Unit.

Several computer packets are available on the research server, such as SAS, SPSS, STATA, GAUSS and R. The programs are frequently updated with new versions.

Rules for Access to Anonymised Micro Data under Statistics Denmark's Research Schemes

At its meeting on 18 June 2008 the Board of Directors laid down the following rules for access to anonymised micro data¹ under Statistics Denmark's research schemes:

Who can get access? Access is only granted to authorised research and analysis environments. Only research and analysis environments of a more permanent nature with a chief researcher and several researchers/analysts can be authorised, as sanctions in case of violation of the rules would otherwise have limited effect. Authorisation is granted by the Director General. The authorisation form is shown in *Appendix 2*.

Prior to granting the authorisation, Statistics Denmark makes a concrete assessment of the applicant's reliability as a data recipient. In respect of non-governmental organisations and enterprises it is relevant to examine the ownership, the staff (qualifications) and the assignments handled for public-sector clients in particular. The examination may include inquiries to such clients to obtain a statement.

When a research or analysis environment has been authorised, agreements may be concluded with specific researchers/analysts. The provisions of these agreements are shown in *Appendix 3*.

The following environments can be authorised:

1. *The user group defined under the framework agreement between Statistics Denmark and the Ministry of Science, Technology and Innovation* can be authorised and thus get access. This group comprises all employees in government funded research projects, employees in public research and analysis environments (i.e., universities, government research institutes, ministries, government agencies, etc.) and researchers employed with non-profit foundations in Denmark.
2. In the private sector, the following organisations with research and analysis environments of a more permanent nature are eligible for authorisation:
 - a. *Non-governmental organisations*
 - b. *Consulting firms* may be authorised, but cannot generally get access to micro data containing business data. The Director General may grant exemption to consulting firms that carry out investigations or research for a public authority, or to a non-governmental organisation that would be eligible for authorisation if its client guarantees, in writing, the correct use of data in terms of security.
 - c. *Other individual enterprises* may be authorised, but cannot get access to micro data containing business data.
3. Danish researchers *who are working abroad for a period*, but who are attached to an authorised Danish research environment, may get external electronic access from their place of research abroad. In these cases, the responsibility lies with the authorised Danish research environment.
4. Foreign researchers may achieve access to micro data under the authorisation of a Danish institution, which assumes the overall responsibility.

What data can be accessed? Access can only be granted to anonymised data, i.e., data for which all identification details such as name, ID number and address have been removed.

¹ There is no access to micro data with identifications.

Access is granted according to the need-to-know principle, which implies that researchers/analysts can get access to the data required for the specified purpose.² Accordingly, the applicants have to document a reasonable relationship between the requested data content and the project description. If the issue requires total population coverage, access may be granted to total data material; otherwise a sample will be made available. In addition, data may be limited in the form of grouping or segments for confidentiality reasons.

Generally, authorised persons have access to all types of personal and business data with the limitations following from the above rules for consulting firms and other individual enterprises, and the need-to-know principle. However, anonymised business data cannot be accessed until one year after the reference period. Detailed product data on individual enterprises are anonymised before they are made available.

In certain cases, Statistics Denmark may deny the requests of a researcher or analyst because of insufficient data quality, primarily in connection with compilation of information from different statistical fields. This applies to both personal data and business data.

Cases of doubt

Acting on the recommendation of the heads of divisions, the Director General decides any cases of doubt resulting from interpretation of the rules.

² This is in accordance with the principles of the Danish Act on Processing of Personal Data, particularly section 5(3): "Data which are to be processed shall be adequate, relevant and not excessive in relation to the purposes for which the data are collected and the purposes for which they are subsequently processed"; and section 10(1): "Data as mentioned in section 7(1) or section 8 may be processed where the processing is carried out for the sole purpose of carrying out statistical or scientific studies of significant social importance and where such processing is necessary in order to carry out these studies."

AUTHORISATION

Statistics Denmark hereby grants

[organisation]

Authorisation for

Remote electronic access to selected datasets at Statistics Denmark

Remote access via the Internet is subject to the following terms:

1. A project description must be submitted, which states the project objectives and renders it possible to select the data required for successful project execution.
2. Based on the project and data description, Statistics Denmark decides whether external electronic access to data can be granted for the specified project.
3. The researcher to whom external electronic access is granted shall sign a special agreement with Statistics Denmark, cf. appendix.
4. All datasets are confidential, cf. subsection 3 of section 27 of the Danish Public Administration Act and section 152 of the Danish Criminal Code.
5. The researcher obtains access to make batch runs on Statistics Denmark's special researcher servers from the research/analysis environment. Access can also via the research/analysis environment be switched to linked-up home computers in accordance with the guidelines determined by Statistics Denmark.
6. Only the client software assigned by Statistics Denmark may be applied in connection with the RSA SecurID card provided. A PC connected to Statistics Denmark may not be made available to unauthorised persons, and when the user leaves the PC, the PC must be either shut down or disconnected, i.e. protected from any unauthorised use.
7. The password of the individual researcher is personal and strictly confidential.
8. The researcher may not, directly or indirectly, download the dataset or any datasets derived there from. All transfers of output for printing or further statistical processing (in spreadsheets or similar) must be executed in accordance with the guidelines and methods laid down by Statistics Denmark. Statistics Denmark will create a log file of such authorised transfers. Furthermore, individual records may not be printed, and all output must be aggregated to an extent that eliminates any risk of direct or indirect identification of persons or enterprises. The researcher may not attempt to make such identification.
9. Statistics Denmark shall be entitled to make unannounced visits to check that the rules of this agreement are observed.
10. The person signing this agreement on behalf of the research/analysis environment shall ensure that publications by the environment do not contain any information that may identify individual persons or individual enterprises.
11. The person signing this agreement on behalf of the research/analysis environment undertakes personally to supervise or to appoint a person to supervise that the provisions of this agreement are observed.
12. In case of breach of the provisions of this agreement, the researcher in breach will be excluded from using any research schemes of Statistics Denmark

permanently or for a period of not less than three years. Furthermore, in the case of breach hereof, this authorisation will be withdrawn for a period.

This agreement, which is signed in two copies, enters into force on *[date]* and may be terminated by either party at three months' notice.

Copenhagen, date

Jørgen Elmeskov

Copenhagen, date

AGREEMENT

between Statistics Denmark and

[Name of researcher]

has been concluded on establishing direct electronic access to selected datasets for the research project:

[Name of project]

cf. description of project dated:

Access is given in accordance with the authorisation granted by Statistics Denmark in connection with special agreement dated: *[Name of institution]*

The agreement is subject to the following terms:

1. The data sets to which access is given shall be treated as confidential information in accordance with Section 27, subsection 3 of the Danish Administration Act and Section 152 of the Danish Penal Code.
2. Processing of the basic data may only be conducted from the research environment for which the authorisation been granted, or access can also via the authorized research/analysis environment be switched to linked-up home computers in accordance with the guidelines determined by Statistics Denmark.
3. A computer linked up to Statistics Denmark may not be placed at the disposal of other persons, and when the computer is not used, the connection shall be completely turned off or disconnected, i.e. protected against unauthorized use.
4. Passwords, which are supplied by Statistics Denmark for the project are strictly personal and shall not be passed on to any third party.
5. Basic data as well as derived data sets shall not, neither directly nor indirectly, be downloaded.
6. All transfers of output (tables, analytical results), etc. for printing or for further statistical processing shall only take place in accordance with the guidelines and methods determined by Statistics Denmark. A logging of these transfers is conducted by Statistics Denmark.
7. Confidential data shall not be printed, including data at the level of individuals or firms, and all output shall be aggregated in such a manner that it is impossible to identify individual persons or individual firms directly or indirectly. Attempts at identifying individual persons or firms are not permissible.
8. Access to the data is given for the period:
9. For the purpose of the agreement the following ip address is used:
10. No information from the project in which it is possible to identify an individual person or individual firm may be published.
11. Published information from the project shall be submitted to Statistics Denmark for scrutiny.
12. The “token”, which has been provided for the project, shall be returned to Statistics Denmark when the agreement expires.

On associated researchers in particular.

13. The responsible person signing the agreement of authorisation for the authorised Danish institution, shall approve and assume the responsibility that all existing rules governing access to micro data are observed by the associated researcher.
14. It shall be the responsibility of the authorised Danish institution to inform the associated researcher of the rules governing the use of micro data, including the rules of confidentiality in force as well as the rules governing downloading of data
15. The associated researcher's access to micro data shall pass through the authorised Danish institution and can also be switched to linked-up home computers in accordance with the rules governing work from home
16. The authorised Danish institution appoints a contact person undertaking the responsibility for all contact with the associated researcher and Statistics Denmark.
17. All invoices concerning the associated researcher are forwarded to and paid by the authorised Danish institution in question in accordance with the terms of invoicing applicable to the institution.

A breach of the provisions of this agreement will imply that access to the data is immediately denied. Furthermore, the person who has signed this agreement will in future be excluded from using any of Statistics Denmark's research schemes. In the case of minor breaches, the person will be excluded from Statistics Denmark's research schemes temporarily for a period of not less than three years.

This agreement may be terminated by either party at 3 months' notice. If the authorisation of the research/analysis environment expires or is changed, this agreement is simultaneously cancelled.

Signatures and date:

Statistics Denmark

Researcher

The researcher's manager, who has signed the agreement of authorisation