

## **Step-by- step practices for national register-based research projects at Statistics Denmark**

Below, you will find a brief – step-by-step overview for the practices and procedures when applying for access to register data through statistics Denmark's research services.

The description is designed to outline the typical main steps seen from a researcher's point of view from the first initial contact with statistics Denmark until the researcher gets access to register data.

### **Step 1: Contact:**

The first step is typically that you contact one of the staff members from the Research Unit at Statistics Denmark for an initial discussion and presentation of the project. The staff members can be contacted by phone +45 3917 3130 or by E-mail. E-mail addresses can be found on the internet. <http://www.dst.dk/en/TilSalg/Forskningsservice.aspx>

### **Step 2: Authorization:**

In order to get access to micro data through Statistics Denmark, it is a pre-requisite that you are affiliated by a research environment pre-approved by Statistics Denmark. In case you are already affiliated with an authorized institution you can move directly to step 2.

In case your institution is not authorized, please contact the Research Service Unit for further information. In general, authorizations can be granted to public research and analysis environments (e.g. universities, ministries, etc.), to research organizations as a part of a non-profit foundation and under certain circumstances companies within the private sector. Statistics Denmark will not grant authorization to single persons.

In order to get an authorization an application needs to be submitted to the Research Service Unit at Statistics Denmark. The application should include 1) a description of the research environment; 2) name and contact information for a responsible manager of the authorization (often the Head of the Department). Statistics Denmark will evaluate the application carefully and make a concrete assessment of the applicant's reliability as a data recipient (e.g.: ownership, educational standard among the staff and the research experience). If the institution is approved for access to micro data an authorization contract between Statistics Denmark and the research institution/environment is signed by the National Chief Statistician – Mr Jørgen Elmeskov and the responsible research manager at the institution (For details, please see appendix A).

The most important point in the authorization contracts is that the person signing the authorization agreement on behalf of the research/analysis environment undertakes personally to supervise that the confidentiality of data is always kept. The researcher may not, directly or indirectly, download the dataset

or any datasets derived there from. 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.

### **Step 3 – Project description and approval**

The next step in the process is to submit a project proposal to Statistics Denmark (1-2 pages, a template can be found in appendix C). It is important to stress that access to data is given according to a so-called “need to know”-principle, implying that you can only get access to the data needed to fulfil your research purpose.

The proposal needs to cover the following paragraphs:

- Description of purpose of the research project
- Population to be studied (Need to know)
- DST registers included in the study (Need to know)
- External data (survey data or data from external registers)
- Register year
- Who will have access to data

In order to assist you in selecting the registers of relevance for your project, you can find comprehensive descriptions of most of the registers provided by Statistics Denmark in the quality declaration compiled for each statistic. The declarations describe quality and content of data as well as information about purpose, history, collection method, accuracy, reference period, and comparability over time. All quality declarations can be found on the internet on the following page: <http://www.dst.dk/en/Statistik/dokumentation/Declarations.aspx>

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 infinite ways, allowing you to produce unique analysis of dynamic processes and fluctuations, using the Danish population. Altogether, data from 250 subject-areas are available for research purposes through Statistics Denmark. The majority of the registers fall into the following headlines:

- |                 |                     |                        |
|-----------------|---------------------|------------------------|
| • Population    | • Social conditions | • Business enterprises |
| • Education     | • Health            | • Housing              |
| • Labour market | • Justice           | • Environment          |
| • Earnings      | • Income            | • External trade       |

If there are external data – taken to mean data not produced by Statistics Denmark - such as survey data or data from other register owners, permission must be granted by the Danish Data Protection Agency. It is your responsibility that the permission is granted.

### **Step 4 – Data extraction protocol**

Based on the project description the Research Service Unit will prepare a detailed data extraction protocol, including detailed lists of variables as well as sets of values, if they are of relevance. Descriptions of all variables can be found in our documentation system TIMES, which is also published on the internet, so far in Danish only.

The 600 mostly used variables have very detailed variable descriptions, including data breaches as well as appendices with special interest. This documenta-

tion is called High Quality Documentation and can be found on the following home page:  
<http://www.dst.dk/da/TilSalg/Forskningsservice/Dokumentation/hoekvalitet/svariable.aspx>

Documentation of other variables produced by Statistics Denmark can be found on the following pages:

*Current variables:*

<http://www.dst.dk/da/Statistik/dokumentation/times.aspx>

*Historical variables:*

<http://www.dst.dk/extranet/staticsites/TIMES3/html/Start000-0000-0000-0000-000000000000.htm>

### **Step 5 – Price calculation and contract is signed**

Based on the data extraction protocol a price will be calculated. At present, DKK 1,229.00 (EURO ~150) is charged per hour for extraction of data (January 2013). 10 hours of free initial advisory services are given for new projects managed by public research and analysis environment. Additional charge for disk storage in the case of large-scale projects and log-on charge for private companies will be charged. Finally, a contract stating price and expected time of delivery is sent to you.

### **Step 6 – Data extraction and working with micro data on the servers**

Data extraction will be initiated shortly after Statistic Denmark has received a signed contract. Data are anonymized by means of a project-specific key before data are entered into the researcher computer (all identifying variables such as CPR numbers, addresses etc. are replaced by project specific random numbers). When research data for a project have been prepared by the Research Service Unit, the data are transferred to the research server where remote access is given via the Internet. Communications via the Internet is protected and encrypted, which secures against unauthorized access.

However, before getting access you have to sign an agreement with Statistics Denmark., cf. the agreement in Appendix B. The agreement 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. It is considered a very serious breach of the agreement between the researcher and Statistics Denmark if you transfer micro data from the research server to your own computer or make attempt to identify individual persons. A breach of these rules may exclude you from using any research schemes of Statistics Denmark for a period of time or permanently. In worst cases, the authorization of your research institution might be withdrawn for a period of time.

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.

# APPENDIX A:

## AUTHORISATION

Statistics Denmark hereby grants

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 machines 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 at 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

# APPENDIX B:

Statistics Denmark

## AGREEMENT

Between Statistics Denmark and **Name** of researcher has been concluded on establishing direct electronic access to selected datasets for research.

Access is given in accordance with the authorisation granted to **(Name of Institution)** by Statistics Denmark in connection with special agreement dated **(Date)**

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 the connection shall be completely turned off or disconnected, when the computer is not used, 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: **Two years with the possibility of extension.**
9. No information from the project in which it is possible to identify an individual person or individual firm may be published.
10. Published information from the project shall be submitted to Statistics Denmark for scrutiny.
11. If a "token", has been provided for the project, it shall be returned to Statistics Denmark when the agreement expires.

On attached researchers in particular.

12. 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.
13. It shall be the responsibility of the authorised Danish institution to inform the attached 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

14. 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
15. The authorised Danish institution appoints a contact person undertaking the responsibility for all contact with the attached researcher and Statistics Denmark.
16. All invoices concerning the attached 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

## APPENDIX C:

Project no. xxxxxx

### Recommendation of Project Approval

#### Authorised institution

Write the name of the institution

#### Project title

Write the project title

#### Project description

Write a short description of the project including the objectives of the project.

#### Population

Define the population(s)

#### Variables included

List the registers and if possibly the variables you want to access

Please also include information on external data or own survey data if such data is part of the project. In such cases please include approval from the Danish Protection Agency.

#### Authorised scientists

Name, Institution, E-mail and Cell phone number of the scientists, whom will be working with the data.

Research Service, Date

Approved for external access

---

Ivan Thaulow