

# **Guidelines and criteria for use of microdata from Statistics Denmark under the authorization of the Department of Political Science<sup>1</sup>**

Peter Munk Christiansen

Drafted by:

Committee on Research Data Management

(Simon Calmar Andersen, Carter Walter Bloch, Mathilde Cecchini,  
Lene Hjøllund, Mette Grønborg Stennicke, Kim Mannemar Sønderskov)

April 2022

## **Introduction**

The Danish population-based registers are an important and unique research tool, enabling researchers to carry out representative population-based studies. The Danish population-based registers will in connection with specific cohorts, intervention studies, and biobanks continue to provide the basis for significant knowledge relevant to the understanding and possible prevention of human diseases and other relevant and desirable goals.

To access these unique data sources, including personal information on the entire Danish population, we must respect legislation and confidentiality while also ensuring flexible access to these valuable resources. Our focus is to ensure that personal data is used in compliance with GDPR<sup>2</sup> and henceforth ISO27001 and ensure a flexible access to these valuable data sources within these constraints.

---

<sup>1</sup> Including The Danish Centre for Studies in Research and Research Policy. The Department of Political Science has authorization number 24, and The Danish Centre for Studies in Research and Research Policy has authorization number 79.

<sup>2</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation - GDPR).

The solution offered by the Department of Political Science consists of VPN access to servers located at Statistics Denmark. The servers contain a number of specific research projects in which the data needed for each project is accessible. Researchers can access projects on the server through a personal computer. Downloading data is neither possible nor permitted, and, thus, all data processing must be performed on the server.

Most datasets on the server contain pseudonymized microdata, i.e., individual-level data on persons (personal data) and data on single companies, firms, or institutions available from national registers or other resources. When conducting research on microdata, you must ensure that no microdata information is transferred to any unauthorized persons. Researchers may access data for the approved research only and must never reveal any microdata information to anyone outside the project. This is by far the most important criterion for any register-based study of individuals, companies, firms, or institutions.

A research project is an independent unit where access to data can be granted upon approval from the relevant authorities. Data must be used only within the boundaries described in the project description. Data from one project cannot be used in another project without approval from the relevant authorities. These typically include the Danish Data Protection Agency, Statistics Denmark, and sometimes the Danish Health Data Authority.

Researchers employed at Aarhus University (CVR no. 31119103) or PhD students enrolled at Aarhus University may gain access to data for which Aarhus University is data controller. Researchers with dual employments (one of these at Aarhus University) may only access personal data during time-periods at which they work on projects related to their employment at Aarhus University. To document employment at Aarhus University, the department may request and store a copy of the letter of employment from Aarhus University. Researchers not employed at Aarhus University need a collaborative agreement and either a data processor agreement or an affiliation agreement (see below).

## **The servers**

Data is stored at servers located at Statistics Denmark. The researcher can access the servers through his or her own computer. Downloading and copying personally identifiable information to

the researcher's own computer is neither possible nor permitted. The server is managed by Statistics Denmark and uses strict security measures that preclude users from downloading information, modifying security settings, and installing and modifying system and software. All personal data has been de-identified; however, according to Danish legislation, data is still considered as personal data. Each server contains specific research projects named with six digits, e.g., 703373, and data is placed in a project according to the permissions of the project.

Each research project works independently of other projects on the servers. One, and only one, research institution is responsible for the project, and each research institution uses a unique Statistics Denmark authorization (cf. footnote 1).

### **Access to the servers**

Please see the document "Forbindelse til forskerservere i Danmarks Statistik" at <https://www.dst.dk/da/TilSalg/Forskningsservice/Vejledninger> for a guide on how to access the remote desktop at Statistics Denmark once you have received your ID and password.

### **Rules, definitions, and advice**

- Access is regulated by Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), Statistics Denmark's rules (see <https://www.dst.dk/da/TilSalg/Forskningsservice/Dataadgang>), and AU's Information security policy and rules version 2.0 or any newer versions. These include (but are not restricted to) the following:
  - Data may not be sought extracted from the server at Statistics Denmark in any way using whatever media. This also includes screen dumps, photographs, manual transcript of the screen, video, FaceTime, Skype, or any other method.
  - When connected to the server at Statistics Denmark, the content of the screen may not be shown to persons who are not themselves granted access to the project.
  - When connected to the server at Statistics Denmark, the computer shall not be passed on to unauthorized persons.
  - The password for accessing the server at Statistics Denmark is strictly personal.

- It is not allowed to access the server at Statistics Denmark from locations where there is any risk that any other third party may unintentionally see the content of the screen (e.g., public areas).
- No attempts must be made to identify individual physical persons.
- New users must sign a contract on data access with Statistics Denmark.

Additionally, the following rules and guidelines apply:

- Violating data security is a very serious breach of the agreement between the researcher and Statistics Denmark. Non-compliance with the terms may exclude a researcher from access to data at Statistics Denmark temporarily or permanently. At worst, Statistics Denmark excludes the entire research environment from the research servers for a period of a few to several months. For that reason, make sure that you understand and comply with the Statistics Denmark “Terms of Agreement”.
- Access All fees regarding data access are to be paid by the researcher's project.
- Publications based on data accessed through the department's authorization must be fully or partially credited the Department of Political Science, Aarhus University (or affiliated centres).
- Only employees at Aarhus University, data processors, and affiliated researchers can access the data.
  - Guidelines regarding access for data processors are described in “Guidelines for the use of data processor agreements with researchers for access to register data at Statistics Denmark” at <https://ps.medarbejdere.au.dk/en/research/data-management>.
  - Guidelines regarding access for research affiliates will be described at a later stage.
- All analyses must be covered by the project description for the project and must be necessary for the project.
- Access to data at the server at Statistics Denmark is allowed only from EU countries and through a personal two-factor VPN access to Aarhus University followed by personal two-factor VPN access to Statistics Denmark.
- Access to data at the server at Statistics Denmark is allowed using a computer with fully operational and updated antivirus software. Aarhus University or the data processor must have implemented procedures ensuring that all computers have updated antivirus protection.

- Users must read and comply with Aarhus University's information security policy (<https://medarbejdere.au.dk/en/informationsecurity>) and these guidelines ([link](#)). Both documents must be read annually.
- Users must inform the department immediately regarding changes in employment for affiliated users.
- Users must inform the head of department immediately in case of the data processor's breach, anticipated breach, or any suspected or actual unauthorized use of sensitive data.
- New users working under the department's authorization may be granted access after supervised training and approval by a department administrator (appointed by the head of department).
- Users are either a data governance expert or refer to a data governance expert (with respect to issues related to the use of microdata from DST).
- Each project must have a (or more) designated data governance expert.
- Data governance experts (DGE):
  - DGE are expected to:
    - Supervise all issues related to accessing personally identifiable information on all projects under the department's authorization and help other users to transfer results etc. from DST's servers to their own computers (see below).
    - Ensure that all analyses are performed in compliance with the permissions for the project and that they are necessary for the project.
    - Ensure that only authorized and relevant, active users have access to a given project (in collaboration with the project leader, see below).
    - to stay informed about the terms of use, to inform project participants about the terms of use, and to supervise project participants in their use of the data.
  - Appointment of DGEs
    - DGE are approved by the head of department after application and after evaluation by the Committee on Research Data Management.
    - The DGE has to be reevaluated in the event of changes in affiliation or if the DGE has not been working with microdata for 2 years. It is the responsibility of the DGE to initiate the reevaluation.
    - Only department employees can be DGE. The DGE's primary affiliation must be the department.

- The DGE must be present at the department on a very regular basis, except for periods of illness and research stays at other institutions.
  - The DGE must have extensive experience working with DST microdata.
  - Preferably, the DGE is the project leader in the projects for which they serve as DGE.
  - Preferably, the DGE is heavily involved (e.g., as coauthor) in the project(s) for which they serve as DGE.
  - A person can only in very special circumstances serve as DGE in more than 3 projects in which the DGE is not project leader/heavily involved.
  - Preferably, the DGE plans to be affiliated with the department for an extended period.
- Project leaders
    - Each project must have a designated project leader employed at the department. The project leader must be listed as the contact person of the project at Statistics Denmark and at the directory of ongoing research projects at Aarhus University.
    - The project leader is particularly obliged to stay informed about the terms of use and to inform project participants about the terms of use
    - It is the project leader's responsibility to find a suitable, approved DGE for the project. Preferably the project leader and the DGE is the same person.
    - In collaboration with the designated DGE the project leader must ensure that only active project participants have access to the Statistics Denmark server and revoke access from inactive participants immediately.
  - Microdata
    - Most data in projects stored at Statistics Denmark consists of microdata, which is data concerning individuals, single companies, firms, or institutions. All microdata must be treated as confidential information and must remain on the secure servers at Statistics Denmark. Even though all identifiers such as for example CPR or CVR numbers have been de-identified (replaced by scrambled identifiers), data is still microdata and may not be transferred out from the server. Even if you delete identifying variables such as the de-identified CPR number, it is still microdata and may not be transferred. If the file you want to transfer contains individual observations, it is NOT allowed – NO MATTER what the variables contain.

- Researchers are obligated to treat all data as confidential information in accordance with the terms and conditions of the Danish Act on Processing of Personal Data.
- Confidential information is defined according to the GDPR, the Danish Health Data Authority, and Statistics Denmark's combined criteria. It applies to any information that relates to less than five identifiable physical persons, companies, firms, institutions, or other units with an identification number (e.g., households or families). This means that tables must contain at least five units per cell and that all statistics must be based on groups of at least five cases. For business statistics, an additional confidentiality rule known as *the dominance criterion* is applied. This implies that if the largest or the two largest enterprises in a table cell showing an economic variable amount to a dominant share, i.e., more than 85% of total revenue, the dominance criterion will subsequently apply, and information is considered confidential. For employment data, the dominance criteria apply to statistics measuring a volume. Here, the dominant share is based on more than 85% of full-time employees.
- Transferring files
  - Only a DGE can transfer results from the Statistics Denmark servers. \$
  - When a non-DGE user wants the DGE to transfer files, the user must send an e-mail to the DGE with the following content:

“I wish to export result file(s) from the Statistics Denmark servers. I am fully informed of the rules governing export of data from Statistics Denmark

(<https://www.dst.dk/da/TilSalg/Forskningsservice/Vejledninger>), the Department of Political Science's guidelines and criteria ([link](#)), GDPR, and the Danish Data Protection Act, and I confirm that the results do neither contain microdata nor individual-level data.

I agree and accept that in case any of the requested files do not comply with the guidelines, my possibility to export files from Statistics Denmark will be closed for a period of three months. Subsequent non-compliance will terminate my access to data for a period of no less than 3 months.”

The e-mail must also contain:

- The name of the folder containing the file(s).
    - The folder must not contain any files that are not intended for transfer.
  - A description of the content of the files.
- 
- Results are information intended for publication, which do not contain microdata.
  - Only aggregated results may be transferred from the secure servers at Statistics Denmark. Files containing aggregated results from analyses can be transferred to the email of a researcher. It is of great importance that the researcher has made sure that the files do not contain microdata information. Statistics Denmark saves all transferred files for six months and randomly conducts inspection of files to make sure that users comply with the rules. If rules are violated, the penalty ranges from a personal warning to a permanent lockout of all users on all the institution's projects at Statistics Denmark. Therefore, make sure that no microdata is transferred. For more details on how to send out results, please consult the guidelines at Statistics Denmark's homepage: <https://www.dst.dk/da/TilSalg/Forskningsservice/hjemtagelse-af-analyseresultater>
  - What CAN be transferred from Statistics Denmark?
    - Results of analyses
    - Aggregated tables that obey the rules
    - Aggregated figures that obey the rules
    - Only aggregated tables, aggregated figures, or aggregated statistics intended for publication in which it is not possible to identify, e.g., individuals, households, families, firms, or other units with an identification number are allowed to be transferred.
    - Tables should have at least five observations in a cell. However, all transfers should be evaluated on an individual basis. Having five observations in a cell does not automatically justify that the information is not personally identifiable. In business statistics, you also need to take into consideration if the two largest companies in a cell account for more than 85% of the total



revenue in the cell. If this is the case, you need to have more observations in this particular cell.

- When transferring multiple tables, make sure that it is NOT possible to identify, e.g., individuals by combining two or more of the tables that you are transferring.
  - Exact medians, minimums, maximums, or percentiles can identify individuals. Make sure to truncate values so they represent five (5) or more individuals. In addition, be aware that it is possible to identify individuals from outliers in a figure. Therefore, be very careful not to have single data points in figures.
  - All output must be manually checked before transferred out. Transfer of uncontrolled output is not allowed and considered a violation of the security rules. Users must know exactly what they are transferring. Statistics Denmark randomly inspects transferred files. If security rules are violated, the penalty ranges from a personal warning to a permanent lockout of the whole research environment  
<https://www.dst.dk/ext/3477468153/0/forskning/Guidelines-for-transferring-aggregated-results-from-Statistics-Denmark--pdf>). See also the section “Sanctions” below.
  - To avoid unintended breaches of confidentiality, transfer of log files, syntax/code, and other types of output are not allowed. The reason for this decision is that the majority of earlier breaches of the security rules were unintended and caused by transferring, e.g., log files with direct or indirect personally identifiable data. If publication of syntax/code is required by a journal, please contact the department's data governance expert about how to proceed.
- What CANNOT be transferred from Statistics Denmark?
    - LOG FILES – since log files may contain error messages or listings with identifiable information.
    - TABLES with less than five (5) observations in a cell.
    - TABLES that, in certain combinations, make it possible to identify individuals etc.

- FIGURES with outliers or single data points.
  - Listings with de-identified variables, such as the personal identification number (PNR). Even if the numbers are de-identified and the file does not contain anything apart from the de-identified number itself, it is still not allowed.
  - Files with microdata information – even if the de-identified variables are removed.
  - Script since it can contain identifiable information. Transfer of script between projects can in some cases be permitted.
- If you discover that the rules set up by Statistics Denmark have been broken unintentionally, please contact the contact person at Statistics Denmark and the head of department immediately.
  - The immediate contact is important since it can be regarded as mitigating circumstances if Statistics Denmark is informed about unintended mistakes as soon as the researcher is aware of the breach.
  - Remember that the same rules also apply if you have sent your own microdata to Statistics Denmark and later wish to work with the data outside the environment at Statistics Denmark. Once the data is located at Statistics Denmark, it cannot leave this environment.
  - If you are in doubt about the rules of transferring specific information from the server, then you should aggregate the output further or contact the project's data governance expert. An unintended violation of the rules can have very serious consequences for you and the entire research environment.
  - Please also consult Statistics Denmark's paper for more information on Data Security <https://www.dst.dk/da/TilSalg/Forskningsservice/Dataadgang> and Statistics Denmark's guidelines regarding transfer of files from the servers <https://www.dst.dk/da/TilSalg/Forskningsservice/hjemtagelse-af-analyseresultater>
- Sending files to the servers:
    - Files with non-personally identifiable information.
    - To place information on the server that do not contain microdata, e.g., syntax/code and documentation, please send the file by email to the contact person at Statistics Denmark.

- The email should contain information about server and project number, where to place the file (complete path, typically in the work data folder), and a statement declaring that the file does not contain any personal identifiable information. Please, send only syntax or documentation when it is necessary and too large to type in manually.

## **Sanctions**

According to ISO27001 (A.7.2.3), a formal and commutable disciplinary process must be in place to take action against users who have committed an information security event. It is the responsibility of each individual user to ensure compliance with criteria and guidelines. It is the responsibility of each data governance expert to ensure training, awareness, and compliance with guidelines for all their associated users.

We have imposed sanctions parallel to those communicated by Statistics Denmark (<https://www.dst.dk/ext/3477468153/0/forskning/Guidelines-for-transferring-aggregated-results-from-Statistics-Denmark-pdf>).

In addition, the first time a project leader or one of the affiliated researchers fail to comply with guidelines resulting in a lockout of the whole institution, there is no additional sanction imposed on the data governance expert.

The second time within five years a project leader or one of the affiliated researchers fail to comply with guidelines resulting in a lockout of the whole institution, the project leader loses the right to download results from Statistics Denmark as well as the right to practice as a project leader for a period of two years.