DANMARKS STATISTIK

December 18, 2015
Research Services

Guidelines for Transferring Aggregated Results from Statistics Denmark's Research Services

Access to microdata at Statistics Denmark's is given to researchers and other analysts, with well-defined projects who wish to analyse data for the benefit of Danish society. It is important that this possibility of access is provided in such a way that both data security and data confidentiality are guaranteed. The individual citizen and business enterprise have a claim that confidential data are treated in the strictest confidence. The general rules governing the treatment of confidential data are laid down partly by the Danish Act on Processing of Personal Data and partly by Statistics Denmark's policy concerning data confidentiality, which can be seen at:
http://www.dst.dk/da/TilSalg/Forskningsservice/Dataadgang.aspx

Data confidentiality is a central precondition for the existence of Statistics Denmark's Research Services. All datasets for which access is granted via Research Services are confidential, cf. subsection 3 of section 27 of the Danish Public Administration Act and section 152 of the Danish Criminal Code. Against this background, researchers are legally obliged – through researcher agreements in writing with Statistics Denmark – only to extract general results and *not* data at individual level of any kind from the research server's located at Statistics Denmark.

Below, a description is given of the main rules for transferring aggregated data from the research server, including what is meant by micro data and detailed tables. Also the procedure in cases of a breach of data confidentiality is described. Finally some advice with respect to avoid breaches of data confidentiality is given.

Which data may be transferred to the researcher's own computer?

It is a general rule that only aggregated results, aggregated tables or figures may be transferred, if identification of individual units is impossible, i.e. neither persons, households, families, business enterprises as well as other units with de-identified serial numbers must be transferred. This applies also if the de-identified serial number is deleted.

In other words, when you transfer data it must be impossible to recognize any individual persons or business enterprises in the data material.

The rule of thumb below generally applies to aggregated data.

The transferred results should be aggregated to a level which can be used directly in a publication.

Further processing of the aggregated data is allowed at the researcher's own computer, e.g. creation of graphs and figures or further analyses in statistical programmes.

The following applies to tables:

Tables should contain at least 3 observations per cell.

The rule of thumb applying to tables is that they should contain at least 3 observations, but the level of aggregation must always depend on a concrete assessment of the content of the table. For further formation, please see chapter 3 of: "Datafortrolighedspolitik i Danmarks Statistik".

Maximum and minimum percentiles, e.g. median may be transferred if and only if there is no risk of identifying individuals. Consequently, it must be impossible to recognize any persons or business enterprises in the material transferred to the researcher's own computer.

Functions such as minimum and maximum frequently refer to individuals, but they can be transferred, if recognition of any persons in the data material is impossible.

Figures

Transfer of figures is only allowed, if they do not contain any identifiable information. With regard to figures, special attention must be paid to outliers and extreme values.

Furthermore, it must be ensured that figures do not contain encased micro data. The latter can be ensured by transferring figures in pdf format.

Which data must not be transferred to the researcher's own computer?

Micro data in any form must not be transferred.

Micro data are defined as data containing, e.g.:

- Datasets or parts of datasets with information at the level of individuals – i.e. where the dataset contains a row for each individual (individual records) and also in cases where the de-identified serial number is deleted, e.g. the personal registration number. That is, a dataset at the level of individuals comprising various items of background information, e.g. income, level of education and socio-economic status, but where the de-identified personal number for each person is deleted.

- The de-identified key variables, such as personal registration numbers, CVR numbers, workplace numbers, address codes, etc. are always considered to be micro data, as they indicate a unique number referring directly to a single individual or enterprise. Even if de-identified person numbers are transferred without any other information, it is still not allowed. This applies also to external data, i.e. data which are sent to Statistics Denmark by the researcher and which are to be incorporated in a specific research project.

Pay attention to program files and log files.

Programs and log files may also contain micro data, e.g.:

- Program code containing micro data written into a condition
- Log files listing individual observations. Some procedures list e.g. a small sample of the dataset that has been used in the analysis.

A transferred log file listing de-identified personal registration numbers is considered to constitute a breach of rules – also if these personal numbers do not contain other items of information.

All work involving processing of micro data must be conducted on the research servers at Statistics Denmark.

It is allowed to construct log files listing de-identified personal registration numbers – e.g. with the aim of checking a program, but such a file must, similar to all other files containing micro data, remain at the research servers and must not be transferred to the researcher's own computer.

Detailed tables should not be transferred from the research server at Statistics Denmark.

Tables containing fewer than 3 observations should not be transferred.

If fewer observations are contained in the tables, a further aggregation of these tables is needed. Alternatively, you can cloud cells with few observations before the tables are transferred. Clouding can e.g. be conducted by deleting the content of the cell or by replacing it with a marking of, e.g. "n<3".

As far as business statistics are concerned, an additional confidentiality rule is applied for economic variables (e.g. turnover or value added) known as the dominance criterion. This implies that if the largest enterprise or the two largest enterprises or kind-of-activity units in a table cell amount to a dominant share, i.e. more than 80 pct. of the value in the cell, the dominance criterion will subsequently apply and the cell must be clouded. For further information, please see chapter 4 of: 'Datafortrolighedspolitik i Danmarks Statistik' [1].

It should be stressed that it must also be impossible from marginal totals, percentages or rates in a table to derive whether the information stems from a too low number of observations. For example, it must be impossible to derive from the aggregated totals in a table, information on the number of persons, who have been subjected to the non-disclosure practice. In other words, it must be impossible to "calculate backwards"

Unlike micro data and with regard to tables, it is impossible to lay down fixed rules as to which data may and may not be transferred. In other words, tables, figures, etc. may be transferred, if there is a higher level of detail than the rules of thumb recommended by Statistics Denmark, provided that identification of single individuals in the output transferred is impossible. For example, it is

[1] Please note that Research Services interprets the value of this criterion more rigorous than describe in 'Datafortrolighedspolitik i Danmarks Statistik'

allowed to transfer the median, although this is, in reality, micro data, provided that identification of any persons or enterprises in the data material is impossible. However, it is important to stress that all transfers of data take place at the researcher's own responsibility.

Procedures and sanctions in the case of breaches of data confidentiality

If there is a breach of data confidentiality, a distinction is made between transfer of detailed tables and transfer of micro data.

If Statistics Denmark observes that micro data have been transferred to the researcher's own computer, the authorized institution's access to data from Statistics Denmark is closed immediately, and Statistics Denmark subsequently complies with the procedure described below.

If Statistics Denmark observes that detailed tables have been transferred, the authorized institution's access to the data is not closed by Statistics Denmark, but we contact the researcher for an explanation. Furthermore, we inform the researcher of our rules governing transfer of data and the matter is subsequently closed. Only in cases of highly detailed tables which resemble micro data, i.e. most cells contain only one observation, Statistics Denmark may subsequently consider treating the matter as transfer of micro data.

Sanctions affecting both the individual researcher and the authorized institution owning a project

Only authorized research institutions can have access to well defined research projects under Statistics Denmark's research arrangement. For Statistics Denmark it is important to have full confidence in the institutions that are granted authorisation. Statistics Denmark has a good knowledge of the authorized institutions, whereas knowledge of the individual researches is far better at the authorized institutions than at Statistics Denmark. Consequently, it is essential that we are fully confident that the institutions can vouch 100% for the researchers, for whom the institutions want to give access to de-identified micro data placed on their research projects. The institutions must guarantee that researchers working on a project have the requisite competences and that researchers are familiar with our data confidentiality rules.

Consequently, at each authorized institution a person in charge is appointed to be responsible for and to supervise – that researchers and analysts attached to their research projects are familiar with Statistics Denmark's data confidentiality rules.

It appears from the signed agreements by Statistics Denmark, the authorized institutions and their researchers that a serious breach of data security or the data confidentiality rules will imply, that researchers breaking these rules can be permanently excluded from Statistics Denmark's Research Services or excluded for a specific period of time. It also applies that the institution owning the specific research project can be excluded from Statistics Denmark's Research Services. Consequently, a breach of rules can result in implications for the person breaking the rules as well as for other research projects and re-

searchers attached to the authorized institution owning the research project in question. The reason for this is that a serious breach of data security or data confidentiality rules indicates that the authorized institution involved has not sufficiently enforced the rules of confidentiality laid down by Statistics Denmark in the research environment. This poses a problem for the Research arrangement in general. Consequently, sanctions are not only directed at the individual researcher but also at the authorized institution responsible for ensuring that the researchers involved are familiar with the rules and that they comply with these rules.

Transfer of detailed tables

In cases where tables are transferred to the researcher's own computer and Statistics Denmark considers that there is a risk of identifying individuals, the researcher who has transferred the file and the responsible person of the authorized institution to which the researcher is attached, is requested by Statistics Denmark to provide an explanation of the transfer of the detailed tables, and to confirm that the files have been deleted. Statistics Denmark also requests the authorized institution to explain how the institution will avoid transfers of detailed tables in the future. In cases of single breaches, there will usually not be any implications for the researcher or the authorized institution as far as access to micro data is concerned. However, in cases of repeated breaches of the data confidentiality rules, where tables have been transferred to the researcher's own computer and there is a risk of identifying individuals, this practice may imply that the institution is excluded from access to micro data. Subsequently, the case will be subjected to further treatment, similar to a case in which micro data have been transferred.

Transfer of micro data

If Statistics Denmark observes that micro data have been transferred by a researcher, the researcher's access to the micro data is immediately closed along with all other projects owned by the authorized institution. If micro data are transferred by a researcher from a project which is not owned by the researcher's own institution, but by another authorized institution, access is closed for the institution owning the research project – but not for the researcher's own institution. Consequently, the responsibility with respect to data security of a researcher rests with the institution owning the project. The access to data for a researcher who has transferred the micro data is closed for all the projects the researcher is connected to, irrespective of which institutions own the projects.

Subsequently, the researcher who has transferred the file and the responsible person of the authorized institution is contacted. They are asked to immediately delete all files breaking the rules governing transfer of micro data and to send a confirmation that all files breaking the rules have been deleted. This applies to both data on hard discs, mail accounts and wherever these data may be kept.

Statistics Denmark will also ask the institution to prepare an explanation of the breach of the data confidentiality rules and the extent of the transferred files breaking the rules. The responsibility for preparing such an account rests with the responsible person of the authorized institution. Subsequently, Statistics Denmark asks the institution to submit a plan describing which measures the institution is going to take in order to avoid breaking the rules in the future. All

accesses to data will remain closed until both the account and the plans are submitted to Statistics Denmark.

Subsequently, Statistics Denmark's management decides which sanctions should be taken towards the institution. Treatment of the case may take up to 8 working days.

The sanction will generally imply that access to micro data for all researchers and projects attached to the authorized institution is closed for a period of at least one month as well as access to all other projects for the researcher involved in a similar period. In cases of repeated breaches, access will be closed for a longer period of time or in particularly serious cases access is closed permanently.

If Statistics Denmark is informed by the researcher that he or she is responsible for having transferred micro data, this will be considered as a mitigating circumstance, and will consequently make it possible to gain access to data again as soon as the nature and extent of the breach are established by Statistics Denmark.

An advisory summary of the sanctions appears from table 1.

Table 1 Advisory summary of sanctions in cases of a breach of the data confidentiality rules or data security.

		Sanctions against the institution			Sanctions against the researcher	
Transfer of data at the level of individuals		**1st time for the institution**	**2nd time for the institution**	**Repeated times for the institution**	**1st time**	**Several times**
a.	Technical mistake which was not done deliberately	Quarantine of 1 month for the institution (may be considered a mitigating circumstance, if the mistake is reported by the authorized institution)	Quarantine of 2 months for the institution	Quarantine of 3 months for the institution	Quarantine of 1 month from all research projects (may be considered a mitigating circumstance, if the mistake is reported by the authorized institution)	Quarantine of 3 months from all research projects
b.	Deliberate action - wanted to look at data (error detection)	Quarantine of 2 months for the institution	Quarantine of 3 months for the institution	Quarantine of 3 months for the institution	Quarantine of 2 months from all research projects	Permanent Exclusion
c.	Deliberate attempts at identifying data	Quarantine of 3 months for the institution	Quarantine of 3 months for the institution	Permanent exclusion	Permanent exclusion	Cannot occur
Password or access passed on by the authorized person		**1st time for the institution**	**2nd time for the institution**	**Repeated times for the institution**	**1st time**	**Several times**
a.	Carelessness	Quarantine of 1 month for the institution	Quarantine of 2 months for the institution	Quarantine of 3 months for the institution	Quarantine of 1 month from all research projects	Quarantine of 3 months from all research projects
b.	Deliberately	Quarantine of 3 months for the institution	Quarantine of 3 months for the institution	Permanent exclusion	Quarantine of 3 months from all research projects	Permanent exclusion

If there is a breach of data confidentiality

If micro data or detailed tables are by mistake transferred, then contact Statistics Denmark immediately.

Always inform Research Services in case of any breaches of rules.

Give a description in your email of when the data were transferred as well as the extent of the data transferred and names of the files. Access to micro data will still be closed, and an account and plan for preventing transfer of data must be submitted to Statistics Denmark, but it is considered to be a mitigating circumstance, if the breach of rules is reported by the researcher or authorized institution at its own initiative.

If Statistics Denmark is immediately informed of any breaches of rules, it is considered to be a mitigating circumstance.

Useful advice

Below, a list of useful advice is given with regard to avoid any breaches of the data confidentiality rules and data security:

For the researcher having access to micro data:

1. Always conduct checks of all files that are to be transferred
    a. In cases of doubt, always contact Statistics Denmark's Research Services

2. Restrict the transfer of files to what is considered absolutely necessary.
    a. Is the transfer, e.g. of programs and log files necessary? – they are kept safely under Statistics Denmark's research arrangement
    b. Is the transfer of provisional results necessary?

For the responsible person of the authorisation

Prepare a plan thus ensuring:

1. That all researchers are informed of the data confidentiality rules and data security laid down by Statistics Denmark
    a. Also ensure that the last ingoing researcher and all foreign researchers have sufficient knowledge of the rules!
    b. That there is a person in the environment to whom contact can be made in cases of doubt

2. That all authorized researchers or analysts for whom you are responsible have signed the research agreement stating that they are familiar with the procedure in case of any breaches of the rules

3. That all authorized researchers are familiar with the ”traps” involved in working with micro data – e.g. that log files can contain micro data

4. If possible, construct control measures when results are transferred.

5. Restrict, if possible, the number of researchers per project who is entitled to transfer outputs (contact Research Services for this)

6. Conduct regular checks to ensure that only relevant researchers have access to micro data for the institution’s research projects and inform Statistics Denmark’s Research Services, if access to data for any researcher is to be de-activated