

## **Procedure for the dissemination of confidential data for scientific purposes**

### **1. General**

- 1.1. This procedure governs the dissemination for scientific purposes of all confidential data available to Statistics Estonia, irrespective of their source, in accordance with section 38 of the Official Statistics Act.
- 1.2. Confidential data collected in the course of statistical activities outside the programme shall be disseminated by Statistics Estonia in accordance with the same procedure.
- 1.3. Statistical activities outside the programme shall include all activities that Statistics Estonia is required to undertake to prepare confidential data and check the output.
- 1.4. Pursuant to section 20 of the Official Statistics Act, there shall be fees for performing statistical activities outside the programme.

### **2. Definitions**

- 2.1. In the sense of the meaning of the Official Statistics Act, confidential data are data that allow direct or indirect identification of a statistical unit and thereby the disclosure of micro-data.
- 2.2. A statistical unit may be directly identified if the data include either the name, address, or personal identification code or registry code of the statistical unit.
- 2.3. A statistical unit may be indirectly identified if, in the absence of characteristics that allow direct identification, it is possible to identify the statistical unit on the basis of other data. In order to determine whether a statistical unit is identifiable, all possibilities that a third person could potentially use to identify the statistical unit shall be taken into account.
- 2.4. A statistical unit is an object or subject described by the data collected, processed and disseminated in the course of producing official statistics.
- 2.5. Micro-data are detailed data characterising a specific statistical unit.
- 2.6. Statistical disclosure control is a set of methods to reduce the risk of identifying individuals, enterprises or other organisations from individual or aggregate data and thereby to reduce the risk of disclosing data about them. The main methods are restricting the amount of disseminated data, e.g. removing from the data the characteristics posing a higher risk of identification, and modifying data, e.g. recoding and micro-aggregation of data.
- 2.7. The risk of identifying a statistical unit indicates the probability of identifying an individual object or subject from the dataset and thereby disclosing the data of the identified unit.
- 2.8. The impact of identifying a statistical unit shows the potential consequences of the disclosure of data for the identified unit, e.g. disclosure of sensitive personal data or business secret, or for Statistics Estonia, e.g. damage to reputation.

### **3. Submitting and processing of the application**

- 3.1. A legal person or institution requiring confidential data for scientific purposes shall submit a written application to Statistics Estonia (example attached). If the application cannot be granted due to insufficient information, Statistics Estonia shall contact the applicant within five working days to specify the data if necessary.
- 3.2. The processing of applications for the dissemination of confidential data for scientific purposes and the advising of applicants is organised by the Marketing and Dissemination Department of Statistics Estonia.

3.3. Applications for the dissemination of confidential data for scientific purposes shall be reviewed in the order of receipt by the Confidentiality Council, convened by an employee of the Marketing and Dissemination Department.

3.4. The Confidentiality Council shall consider the grounds for the application and make a decision about granting the use confidential data for scientific purposes within ten working days of receiving all the documents necessary for the evaluation of the application.

3.5. The Confidentiality Council shall consider each requested dataset from two aspects: the risk of identifying a statistical unit and the impact of identifying a statistical unit.

3.6. An employee of the Marketing and Dissemination Department of Statistics Estonia shall inform the applicant of the decision in writing.

3.7. If the application is granted, Statistics Estonia shall enter into an agreement with the legal person or institution that submitted the application before disseminating the data. Pursuant to subsection 38(2) of the Official Statistics Act, such agreement shall not be entered into with a natural person.

3.8. Statistics Estonia shall not enter into an agreement with a legal person or institution that has violated the terms and conditions of an agreement to transmit confidential data.

#### **4. Confidentiality Council**

4.1. The members of the Confidentiality Council shall be approved by a regulation of the Director General of Statistics Estonia.

4.2. The Confidentiality Council shall have a quorum if all the members of the Council are present or at most one member is not present. Decisions shall be taken by a simple majority. Dissenting opinions shall be recorded in the minutes. If the votes cast for and against are equally divided, the Chairman of the Council shall have the casting vote.

4.3. The meeting of the Confidentiality Council may be conducted electronically.

4.4. During an absence from work, a member of the Confidentiality Council shall be replaced by an alternate appointed.

#### **5. Statistical processing, dissemination and storing of confidential data**

5.1. Statistics Estonia shall prepare the data and metadata disseminated for scientific purposes.

5.2. The data shall be disseminated as a text file (CSV) where one row contains data for one statistical unit. There shall be a separate data file for each survey year. The data shall include the survey code, survey year, unique code of the statistical unit and variables collected and calculated in the course of the survey.

5.3. The metadata to be disseminated shall include a description of the variables (name of variable in the database, explanation of content, variable type, range of values, classification or codifier used, calculation for a derived variable), description of the survey methodology and questionnaires used to collect the data. Metadata shall be made available together with the data.

5.4. In the case of a high risk of identification and/or high sensitivity, the use of confidential data may be permitted only at a local workplace. The Confidentiality Council shall have the authority to make the relevant decision.

5.5. The members of the microdata team of Statistics Estonia shall apply statistical disclosure control to the whole output that the user wishes to receive to an e-mail address.

5.6. If the output contains indirectly identifiable data, the first time, a warning shall be given to the user. The next time, the institution's contact person shall be informed.

5.7. Statistics Estonia shall store for 25 years all confidential data, disseminated to users for scientific purposes, in the form that they were made available for the specific research work. Statistics Estonia shall store log files, users' interim results and other log files for 1 year after the expiry of the agreement. If the user wishes that the interim results or work files were kept for more than a year, the user shall cover the cost of storing these.

## **6. Local workstation**

6.1. Local workstations are in the offices of Statistics Estonia, located at Tatari 51 in Tallinn and at Narva mnt 20 in Tartu.

6.2. Local workstations may be used on the basis of pre-registration on working days from 9:00 to 16:00 or remotely via VPN around the clock.

6.3. On a local workstation computer, each user shall have for each research work a separate folder (20 GB) where the user can store interim results of the work and where back up shall be ensured by Statistics Estonia. Each user shall have access only to the datasets necessary for the user's research work.

6.4. On a local workstation computer, the user shall not have the possibility to print, copy files to external media or use the internet.

6.5. If the user needs to save files on a local workstation computer for the purposes of research, an employee of the Marketing and Dissemination Department of Statistics Estonia shall save the files.

6.6. After a disclosure control performed by Statistics Estonia, a member of the microdata team of Statistics Estonia shall send the research output to the user's e-mail address.

## **7. Checking compliance with the procedure**

7.1. Compliance with the procedure shall be checked by the Marketing and Dissemination Department of Statistics Estonia.

## **Example of an application for the use of confidential data**

An application shall contain the following information:

1. name of the legal person or institution;
2. registry code of the legal person or institution;
3. title of the research work;
4. objective of the research work;
5. reference to the research work according to the Estonian Research Information System (if any);
6. name of the statistical activity or a list of the data required for the research;
7. list of the data that the applicant has obtained from other sources and that the applicant wishes to link with the data applied for;
8. list of persons wishing to use the relevant data in the course of the research (name, personal identification code, e-mail address);
9. name and job title of the signatory of the agreement;
10. how the research output is made available to the public (article, monograph, etc.);
11. period of using the researchers' environment (for what period are the data needed);
12. term of using the data, i.e. the term of agreement;
13. format of the requested data (if differs from CSV);
14. if necessary, a request to use own source codes on the local workstation.

(Signature)

Name and job title of the person with the right of representation

Name and address of the legal person or institution