



Economic and Social Council

Distr.: General
6 June 2013

English only

Economic Commission for Europe

Conference of European Statisticians

Sixty-first plenary session

Geneva, 10-12 June 2013

Item 4 (a) of the provisional agenda

Drivers for micro-data access

Statistics Estonia's experience in providing national and trans-border access to micro-data

Note by Statistics Estonia

Summary

The aim of the paper is to share the Estonian experience in the adoption of new practices in micro-data exchange, supported by the experience in developing new statistical law. This paper consists of a description of the policy and the implications of innovation in access to micro-data. Other changes in the Estonian law on official statistics are not covered.

I. Introduction

1. The representatives of Statistics Estonia were included in the inter-institutional working group on access to micro-data for research purposes. To develop regulations to the access to micro-data, the group studied the corresponding practice of other countries. Based on information gained at international meetings, at conferences, from the websites of statistical offices, but also from the UN/ECE/CES handbook “Managing Statistical Confidentiality & Micro-data Access: Principles and Guidelines of Good practice”¹, the relevant chapter of the Act was drafted. It was decided that the detailed procedures will be described in a separate document to be approved by the Director General of Statistics Estonia. Also, certain institutions were visited in order to study their experience and practices more carefully.

2. In 2010 the representatives of the Ministry of Finance, the Ministry of Education and Research, Eesti Pank (the Bank of Estonia) and Statistics Estonia made study visits to four institutions: Luxembourg Income Study, Statistics Finland, Statistics Denmark and Statistics Netherlands.

3. During the study visits it became clear that it would be possible to base on necessary documentation (e.g. application form, oath of confidentiality, agreement with research institution, etc.) used by other statistical offices, modify it and shortly introduce it at Statistics Estonia. It also appeared that the LISSY software, developed and used by the staff of Luxembourg Income Study, was suitable for the provision of remote execution service to Statistics Estonia’s customers.

II. Current legislation and practice

4. In August 2010, the new law on official statistics came into force in Estonia. By the end of 2010, the Procedure for access to confidential data for scientific purposes had also been approved. Access to micro-data is currently governed by sections 35–38 of the Official Statistics Act² and the Procedure for dissemination of confidential data for scientific purposes³.

5. All categories of micro-data collected for the production of official statistics (social survey data, census data and business data) are available for scientific purposes without the consent of the person, i.e. micro-data of both natural persons and all kinds of economic entities. Also, data derived from administrative records and other databases may be accessed for scientific purposes.

6. Data permitting direct identification and data permitting indirect identification of a statistical unit are both allowed to be accessed for scientific purposes.

7. Micro-data are allowed to be accessed for scientific purposes only by legal persons or agencies, but not by freelance natural persons. Pursuant to the Estonian law, “a research

¹ Managing Statistical Confidentiality & Micro-data Access: Principles and Guidelines of Good practice. (2007). UN.
http://www.unece.org/fileadmin/DAM/stats/publications/Managing_statistical_confidentiality_and_micro-data_access.pdf

² Official Statistics Act.
<http://www.legaltext.ee/et/andmebaas/tekst.asp?loc=text&dok=XXXXX42K1&keel=en&pg=1&ptyyp=RT&tyyp=X&query=riikliku+statistika>

³ Procedure for dissemination of confidential data for scientific purposes.
<http://www.stat.ee/dokumendid/51669>

and development institution” is an institution specified in section 3 of the Organisation of Research and Development Act, or a university or another establishment providing higher education of a foreign state or a research institution of a foreign state, or an institution listed in the relevant Decision of the European Commission. Pursuant to section 3 of the Organisation of Research and Development Act, a research and development institution is an institution in case of which: the principal activity is carrying out basic research, applied research or development; the activity accompanying principal activity is to spread knowledge through teaching, publication or technology transfer; the results of the principal activity financed from the state budget funds (which do not involve intellectual property rights) are public information; the membership includes the research staff necessary for carrying out the principal activity.

8. Students pursuing a Master’s or Doctor’s degree are also considered researchers. The same rules apply to domestic and foreign research institutions.

9. A legal person or agency in need of confidential data for scientific purposes must submit a written application to Statistics Estonia. The application must set out the following information: name of legal person or agency; registration code of legal person or agency; title of the research; objective of the research; name of the statistical action or a list of data necessary for the research; a list of data which the applicant has obtained from other sources and which the applicant wishes to link with the data applied for; a list of persons wishing to use the relevant data during the research (given and surname, personal identification code, email address); in case the use of personal data is involved, a confirmation issued by the Data Protection Inspectorate to prove that the applied organisational, physical and information technology related security measures are sufficient and, if an ethics committee has been founded, also the opinion of such committee; in case the use of sensitive personal data is involved, a confirmation issued by the Data Protection Inspectorate proving that the processing of sensitive personal data has been registered.

III. Risk and impact based decision procedure

10. Statistics Estonia has the obligation to consider each application separately. Applications for the access to confidential data for scientific purposes are reviewed by the Confidentiality Council according to their order of arrival. The Confidentiality Council considers the substance of the application and decides whether the confidential data can be used for scientific purposes, whereas the decision must be made within ten working days from the receipt of all documents necessary for evaluation of the application.

11. The Confidentiality Council consists of public servants of Statistics Estonia: mainly the heads of statistical departments, but also the head of information services, the person responsible for personal data protection, and the lawyer.

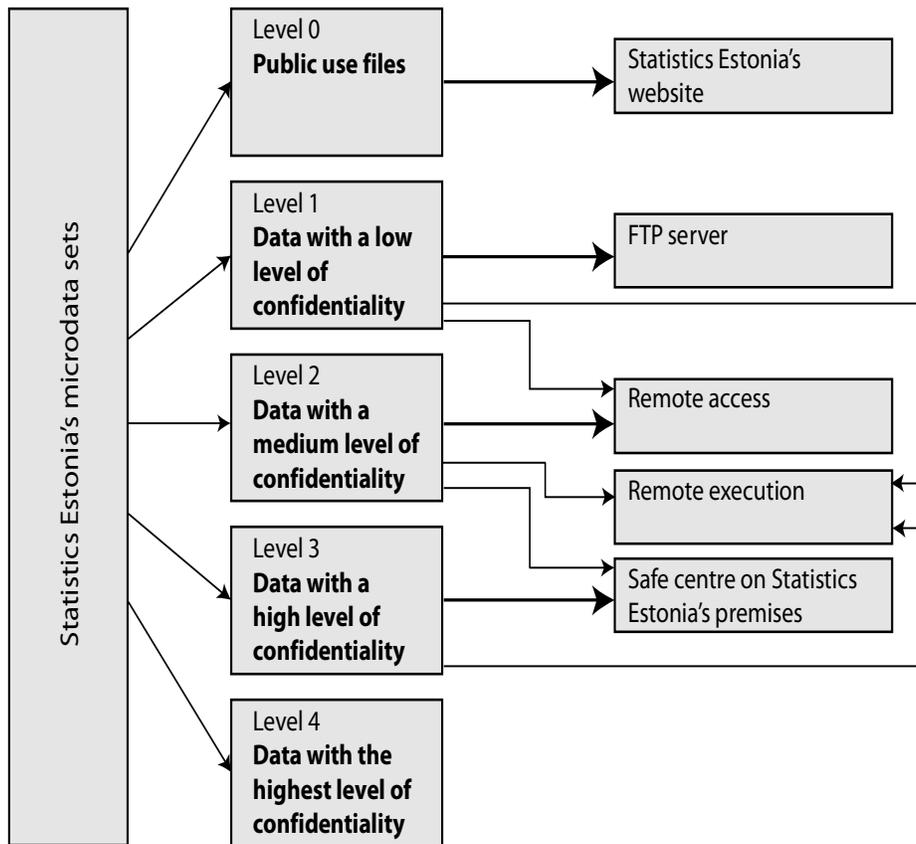
12. The Confidentiality Council meets once a week, if an application has been received. Decisions taken by the Confidentiality Council shall be confirmed by a directive of the Director General.

13. The Confidentiality Council shall consider every dataset requested by the applicant, taking into account two aspects: 1) risk of identifying a statistical unit; 2) the impact that the identification of a statistical unit may have (sensitivity of data). If the identification risk and sensitivity are rated as high, confidential data can be used only in a safe centre, by remote access or remote execution. If both aspects are rated as low, the requested data can also be delivered on removable devices.

14. In case of Statistics Estonia, the following means for access to micro-data are available:

- (a) Remote access;
- (b) Remote execution;
- (c) Safe centre;
- (d) Scientific use file on CD-ROM or via FTP server;
- (e) Public use file on the web;
- (f) Order for information.

15. The following figure outlines the system of access to micro-data at Statistics Estonia according to the level of confidentiality of micro-data.



16. Remote execution means a service whereby users can use micro-data for making statistical analyses in a manner which precludes the users' direct access to micro-data, i.e. a researcher submits a query and receives the output later over the Internet. Remote access means a service whereby a researcher performs the analysis and can immediately see the answer on the screen. Safe centres are located in the offices of Statistics Estonia in Tallinn and in Tartu. In case of all these three modes of access, the results are sent back via computer networks after checking for confidentiality. Scientific use file means micro-data to which methods of statistical disclosure control have been applied, in order to reduce the risk of identification of the statistical unit to the appropriate level in accordance with current best practice. Public use files are published on the website of Statistics Estonia and do not allow any direct or indirect identification of a statistical unit, because these data files

are prepared by applying statistical disclosure control methods. Orders for information mean that the staff of Statistics Estonia produces tailor-made information for a particular user.

17. Statistics Estonia may refuse to disseminate micro-data only if:

- (a) It is not convinced that the data will be used only for scientific purposes;
- (b) The applicant wants to use personal data but the conditions provided for in the Personal Data Protection Act are not fulfilled;
- (c) The applicant has previously violated the terms and conditions of an agreement entered into with Statistics Estonia and, in the opinion of the producer of official statistics, the applicant has not implemented sufficient measures to prevent violation of the terms and conditions of an agreement in the future.

18. Before the release of confidential personal data for scientific purposes, Statistics Estonia shall enter into an agreement with the user of micro-data stating the purpose of the research, the persons entitled to access to micro-data, the procedure for processing or transmission of data and the obligation to ensure the organisational, physical and information technology related protection of data, and conditions for the destruction of data after completion of the research.

19. All users of micro-data mentioned in an agreement shall sign a written confidentiality agreement.

20. The use of micro-data is free of charge. Readily available datasets are listed in Appendix 1. A user shall only pay for the tailoring of the data for their needs and purposes (i.e. linking, matching, etc.): the cost is 100 euros for the first set of data per year and 50 euros for every subsequent dataset.

21. Penalties for the violation of relevant regulations are stipulated in section 40 of the Official Statistics Act. The unlawful dissemination of data which have been collected during the production of official statistics or which enable the identification of a respondent, or the use of data for other than statistical purposes is punishable by a fine of up to 200 fine units (800 euros). The same act, if committed by a legal person, is punishable by a fine of up to 3,200 euros.

IV. Supply has created demand

22. Since the adoption of the current policy on 19 October 2010, Statistics Estonia has received and processed 74 applications (as of 23 April 2013) leading to the conclusion of 44 agreements on remote access or use of safe centre, and 15 agreements on FTP services. In case of 4 applications, an agreement is being prepared. 7 applications have been cancelled and 4 have been rejected.

23. Mostly applications are from Estonian universities, research centres and research departments of public agencies, but a number of applications are received also from foreign universities and international organisations. All conditions for use of confidential data for scientific purposes are the same for researchers from Estonia and outside. All metadata and technical guidelines concerning the use of micro-data are also available in English for better serving foreign researchers.

24. Usually foreign researchers use data via remote access which is especially suitable for outside researchers as they cannot come to use data in the premises of Statistics Estonia. Two of the agreements have been concluded between Statistics Estonia and the OECD, one with University of Essex, London School of Economics and Political Science and Södertörn

University (Sweden) on access to micro-data using remote access. Exceptionally the LIS (Cross-National Data Center in Luxembourg) received data of Estonian social survey using ftp-server. Exception was made by Confidentiality Council taking into account the purpose to update Luxembourg Income Study database.

25. In April 2013 Statistics Estonia conducted a satisfaction survey of researchers mentioned in agreements for remote access or use of safe centre (both past and current agreements). Respondents assessed various aspects on a 10-point scale: satisfaction with the application procedure, speed of access to micro-data, instructions on using the facility, quality of micro-data, selection of available software, user support offered by Statistics Estonia, speed of receiving research results and IT resources of the workstation. The overall level of satisfaction can be considered excellent as the net promoter index was 68%.

V. Linkage of data for external research purposes

26. Linkage of data from different sources gives researchers the possibility to carry out enhanced research based on a wider range of variables than is available in a single data source. For production of new composite datasets for research purposes, record linkage has been used at Statistics Estonia. Several preconditions have to be met for data linking; these are listed in the “Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical or Related Research Purposes”⁴.

27. Principle 1 states that data linking should be undertaken by NSOs only for statistical and related research purposes. To comply with this principle, all research databases are kept in the server of Statistics Estonia. Direct identifiers (personal ID, enterprise register number) of statistical units are recoded with anonymous codes and the key between initial identifiers and the code is kept by Statistics Estonia. The procedure is the same for statistical databases, administrative sources and researchers’ own data. Data linking can be performed by the researcher or by the staff of the NSO using anonymous codes. Both practices are used depending on the research request.

28. Principle 3 states that the public benefits of any data integration project should be sufficient to outweigh any privacy or confidentiality concerns. The public benefits of data linking should be approved by publicly available research output. In the research application, researchers specify the objectives of the research. The agreement on the use of confidential data for scientific purposes⁵ states that the researcher shall make the results of the research (article, monograph, etc.) available to Statistics Estonia or, in case a limitation has been imposed on the publication of the results of the research, a reference indicating the publication thereof in a scientific journal or other source. Links to the research results are published on the website of Statistics Estonia.⁶

29. An excellent example of the benefits of data linking is the study “The Role of Product Level Entry and Exit in Export and Productivity Growth: Evidence from Estonia”⁷. This research used detailed foreign trade data about the full population of exporting firms in Estonia, covering the period 1995 to 2009. Statistics Estonia provided foreign trade data

⁴ Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical or Related Research Purposes. (2009). UN.
http://www.unece.org/fileadmin/DAM/stats/publications/Confidentiality_aspects_data_integration.pdf

⁵ Agreement on the Use of Confidential Data for Scientific Purposes.
<http://www.stat.ee/dokumendid/55077>

⁶ <http://www.stat.ee/teadustood>

⁷ Research working papers of the Faculty of Economics, University of Tartu, No 86, 2012
<http://www.mtk.ut.ee/orb.aw/class=file/action=preview/id=1129916/Febawb86.pdf>

which was integrated with custom data and intra-EU foreign trade survey data. The detailed export dataset has been merged with firm level information about performance indicators and other firm level controls (such as size, age, etc.). This firm level information for the entire population was based on the database of the Estonian Central Commercial Register. The matching of the two datasets was executed by researchers based on the anonymous codes for firms. The Commercial Register's firm level database included annual reports with balance sheets and profit and loss statements for all Estonian firms. This data was employed to calculate the productivity of firms, and to calculate some control variables for the regression analysis. Based on the linked database, the researchers could show that starting export activities and adding and dropping export products in the same period are associated with higher firm productivity compared to exporters that keep their export mix unchanged or decrease its breadth.

30. Principle 6 states that the number of unit records and data variables to be included in a linked dataset should be no more than required to support the approved purpose(s). The list of variables included in a linked database is agreed with researchers and added as an appendix to the Agreement on the Use of Confidential Data for Scientific Purposes. In practice, the staff of Statistics Estonia has suggested to add or remove variables depending on the objective of the research. Any request for additional variables after the agreement has been signed must be approved by the Confidentiality Council.

VI. Conclusions

31. As a result of public debate, the new law on official statistics is quite detailed, including the regulations concerning micro-data. The Official Statistics Act includes 62 sections, instead of the fourteen of the previous version.

32. There is no point in reinventing the wheel. By using the best practices already available in statistical offices, a relevant and up-to-date policy for access to micro-data can be implemented relatively easily and quickly.

33. Statistics Estonia has observed that supply creates demand; also, a clear and quick procedure enhances access to micro-data.

Appendix 1

Micro-data available for researchers

The following data files are available for research use and linking. Additional data files are available on request:

Data files of population and social surveys

Population and Housing Census, 2000

Migration

Estonian Social Survey

Labour Force Survey

Household Budget Survey

Household Module Survey

Time Use Survey

Data files of enterprise surveys

Statistics on the economic activity of enterprises

Structure of Earnings Survey

Foreign trade

Innovation

Agricultural census, 2001, 2010

Data files from administrative sources

Social tax data
