

**Joint ECE/Eurostat Work Session on
Statistical Data Confidentiality**

(Skopje, The former Yugoslav Republic of Macedonia,
14-16 March 2001)

Working Paper No. 31
English only

Topic IV: Progress in the implementation of SDC methods and techniques in central and eastern Europe

**DATA PROTECTION MEASURES IN ESTONIAN POPULATION AND
HOUSING CENSUS OF 2000**

Invited paper

Submitted by the Statistical Office of Estonia ¹

I. INTRODUCTION

1. This report will give an overview of how Census data protection is organised and what kind of protection measures are applied in practice in the Statistical Office of Estonia (SOE). The paper deals with the entry and coding of the data collected by the Census.

II. LEGISLATION

2. In Estonia, the protection of confidential statistical data is based on several legal acts and regulations. The basic provision concerning personal data is contained in the Constitution. The Official Statistics Act was passed by Parliament in June 1997 and amended in 2000. This Act establishes the legal basis for the conduct of official statistical surveys. Official statistics should conform to the principles of impartiality, reliability, relevance, cost-effectiveness, confidentiality and transparency. Statistical disclosure control (SDC) issues are legally regulated by the Personal Data Protection Act passed in 1996, the Databases Act passed in 1997, the Official Statistics Act and the Population and Housing Census Act, passed by Parliament in May 1998, where rules to protect census data have been fixed. Resulting from the amendment to the Official Statistics Act of 2000 the procedure for protection of the data collected and processed by the SOE has been established by a Government Regulation on 29 January 2001.

3. Lower-level norms such as regulations, decrees, official decisions provide rules for the application of the related laws. For the SOE, the protection of statistical data is very important and today more and more attention is paid to its legal and administrative aspects. This does not mean that the other aspects, such as organisational and computing ones, are less significant.

III. DESCRIPTION OF CENSUS DATA PROCESSING

4. The whole Population Census has been a long process having lasted about 10 years, from 1993 through to 2002. The data entry and the first stage of coding of the collected data took about five months, from May through October 2000.

5. After the Census about 6,500 Census portfolios were taken to the Statistical Office. The portfolios contained notebooks of enumerators, forms (so-called accompanying notes) filled in by enumerators, their supervisors and Census district heads, and Census questionnaires (3.5 million A4-format two-sided filled-in sheets), of which 2 times 1.4 million questionnaires carrying personal data and 0.7 million questionnaires carrying housing data.

¹ Prepared by Mati Sundja (e-mail: mati.sundja@stat.ee) and Tiiu Vallner (e-mail: tiiu.vallner@stat.ee).

6. The Census questionnaires carried sensitive personal data because people were asked about their state of health (long-term illness or disability), religious affiliation and also ethnic nationality.
7. Data processing was performed in the main building of the SOE (so-called data processing rooms) that were prepared (underwent repairs, etc.) specifically for this purpose. The data were entered and coded by the 135 computer operators temporarily employed by the SOE. The work was done on working days, in two shifts, using two powerful scanners. Fifteen employees were engaged in optical character recognition and 50, in data coding. In addition to the administration, a security officer, shift managers, clerks of the record, data and system managers, specialists in methodology and advisers, and support persons of contract partners were also involved in data processing.
8. The Census questionnaires were scanned using two high-volume Fujitsu scanners and interpreted using the Eyes&Hands software. During interpretation, operators resolve the characters or words that cannot be recognized by the OCR/ICR software. The output of interpretation is text files.
9. The text files will be coded using the Oracle-based software. If fields in the text files do not match with the pre-defined data dictionaries, operators will be asked to resolve the situation. More complex cases will be forwarded to chief operators to handle. The process is self-learning, i.e. typical answers will be added to dictionaries automatically. In addition, context verification rules will be applied (i.e. child should not be older than parent(s), etc.). The whole process will be monitored from a security aspect, i.e. all data access events will be logged and suspicious transactions may be cancelled any moment.
10. The full database will be archived and stored in a highly secure archive. At the next stage, primary and secondary person identification information will be removed. As a result, an anonymous Census database will be created. The anonymous database will be processed by dedicated software for generation of tabulations and providing access to the database. The anonymous database can be used with the Oracle Reports, Oracle Discoverer and a number of other tools, including the GIS software. The Census data processing software is powered by the Sun Ultra Enterprise 450 server and the Oracle version 8.0.5.

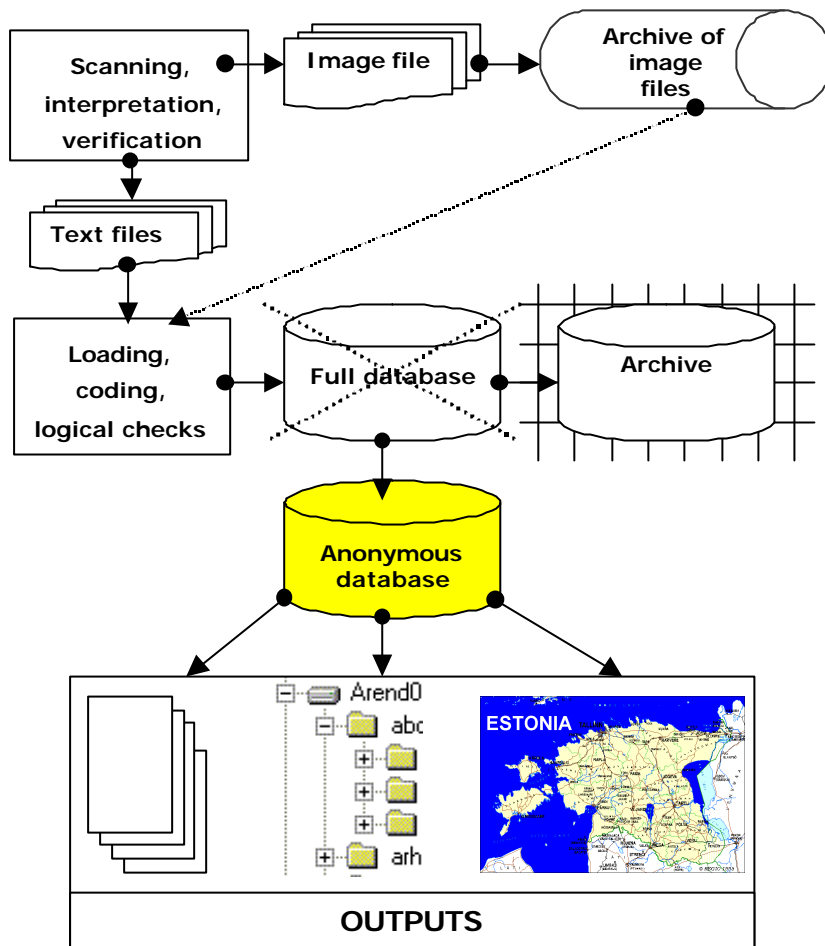


Figure 1. Simplified diagram of Census data processing

IV. LEGAL PROVISIONS FOR THE PROCESSING OF PERSONAL DATA

11. Pursuant to the Personal Data Protection Act, the processing of personal data has to be registered at the Data Protection Inspection (DPI) and organisational as well as technical measures have to be taken to protect the data. The nine items (obligations) of Section 2 of Article 12 of the Personal Data Protection Act that served as a basis by the DPI in checking the security measures taken are to be fulfilled.

In the automatic processing of personal data, chief processors and authorised processors are required to:

- i) prevent access of unauthorised persons to equipment used for processing personal data (access control);
- ii) prevent the unauthorised reading, copying, alteration or removal of data carriers (data carrier use control);
- iii) prevent the unauthorised recording of personal data and alteration of recorded personal data (recording control) and to ensure that it be subsequently possible to determine when, by whom and which personal data were altered;
- iv) prevent the unauthorised use of a data processing system for the transmission of personal data by data communication equipment (data communication control);
- v) ensure that every user of a data processing system has access only to the personal data authorised to be processed by him or her (access control);
- vi) store information concerning disclosure of personal data regarding when, to whom, by whom and which personal data were disclosed (disclosure control);
- vii) ensure that it be subsequently possible to determine when, by whom and which personal data were entered into the data processing system (input control);

- viii) ensure that the unauthorised reading, copying, alteration or erasure is not carried out in the transmission of personal data by data communication equipment and in the transportation of data carriers (transport control);
- ix) organise the work of enterprises, agencies and organisations in a manner that allows compliance with special data protection requirements (organisational control).

12. To ensure the protection of data, security needs are to be determined, security requirements specified and the security category is to be established. Data security is divided into three classes:

- integrity ;
- availability;
- confidentiality.

13. Once the security category has been determined, to each security category element standard measures are applied. The security measures, in their turn, are divided by type (organisational, infrastructure-related, staff-related, soft- and hardware-related, communications, service continuity) and effect (preventive, discovering and reproductive). Thereafter the existing and necessary security measures are determined, the cost and effectiveness of application of security measures are assessed, the action plan is drawn, etc. As there are too many types of security measures, they have a number of parameters and the effects of measures overlap, then special programmes are being developed to plan them. It will be possible to apply such a methodology in the near future. At the SOE, no particular security categories have yet been determined, reference methodology and special computer programmes were not used.

V. PRACTICAL STEPS IN CENSUS DATA PROTECTION

14. The SOE started to work out the Population Census security policy already in 1997 in co-operation with acknowledged companies and continued this work after the Pilot Census. In 1999, a new version of the Population Census security policy was drafted. The report gave an overview of the assets related to the Population Census infosystem. Analysis of the risks associated with the Population Census infotechnological (IT) system, including danger to the Census materials, data in a digital form, hardware and equipment, software, rooms, staff, communication services, was presented, and the degree of the risks was assessed.

15. The major risks included:

- trouble in the availability of Census materials, of which destruction to a large extent;
- protection of the confidentiality of questionnaires carrying personal data, including personal database, i.e. getting of the whole database into the hands of persons not concerned;
- trouble and destruction of IT equipment, data included, that causes non-timeliness of the processing of Census data, double work, etc.

16. The Population Census security policy was worked out proceeding from the recommendations of the EVS ISO/IEC 13335 standard according to which security policy specifies general aims and actions to be planned in the following fields:

- aims and principles of security;
- organisation and infrastructure of security;
- analysis of security and risks of data processing, and risk management strategy;
- information sensitiveness and risks;
- hard- and software security;
- communication security;
- physical security;
- staff security;
- documents and data carriers security;
- planning of after-damage measures and continuous service.

The renewed version of security policy served as a basis for working out a general approach to and requirements for Population Census security measures. On this basis the security rules of Population Census data processing were documented at the SOE.

17. The Population Census security rules cover all the security measures applied to data processing by the above field. The general part of the security rules contains references to the Population Census-related legal acts and the procedures and instructions existing at the SOE. The instructions and procedures that ensure Population Census data processing security and were documented as annexes to the security rules are the main part of the latter:

- 1) IT system security rules
- 2) Physical security requirements
- 3) Procedure for regulation of access to the IT system
- 4) Rules for software procurement and use
- 5) Procedure for IT equipment account
- 6) Storage procedure
- 7) Rules for rebuffering parasitic software
- 8) Rules of procedure of trouble management and reproduction
- 9) Plan of action in a dangerous situation.

18. In addition, the bylaws of the Population Census Division and security requirements in job descriptions of the Population Division's officials and computer operators were approved and security requirements were added to the contracts concluded with partners.

VI. SECURITY MEASURES APPLIED

19. The following security measures were applied:

- IT security measures to protect data against the danger arising out of infotechnological means (Annexes 1, 3 and 7 to the Security Rules);
- physical security measures to protect data from the danger arising from the environment and the persons not concerned (Annex 2 to the Security Rules);
- organisational measures, i.e. organisation of labour that diminishes the danger arising out of the activities of people (bylaws, Annexes 3 to 9 to the Security Rules, job descriptions of employees, security requirements to be met by specialists of contractors).

VI.1 Infotechnological security measures

20. IT security measures were the following:

- data processing was performed in a local network, i.e. there was no physical connection with either the whole network of the Statistical Office or the external network, the Internet included;
- data processing software was worked out in cooperation with specialists of an acknowledged company;
- data processing software is based on the Oracle database management system which has vast possibilities for data protection, for example, different user rights, access, passwords, activity monitoring, etc.;
- a special programme, the so-called security diary, registering all user activities was worked out;
- data processing software enables data crypting that is used to store data allowing identification of individuals;
- possibility of saving and monitoring systemic logs;
- spare copies were made regularly (each day) and stored in safes in accordance with the storage procedure; spare copies of the RL-2000 (Oracle base) database were checked for reproduction;
- routine maintenance of IT equipment was performed by specialists of contractors;
- in the IT system reports on the main jobs of each shift day were drawn;
- list of IT equipment with technical specifications was compiled (procedure for IT equipment account).

VI.2 Physical security measures

21. Physical security measures were the following:

- data processing rooms underwent repairs;
- computer operator jobs were manned;

- air conditioners were mounted, windows were equipped with blinds;
- security doors and locks, including emergency locks, code locks, locks to be opened by a magnetic card, were mounted;
- a card system granting different persons a right to open and lock doors and enabling movement monitoring was introduced;
- alarms and watch cameras were mounted and connected, a video telephone was mounted at the entry;
- fire-extinguishers of different capacities and possibilities, including extra high-power gas-extinguishers for the document deposits, were procured;
- extra sensors (water, temperature, smoke, glass breaking sensors, etc.) were mounted in document deposits and server rooms by way of additional measures;
- safes were purchased;
- in case of electricity supply trouble emergency feeders were used (a separate UPS for equipment in data processing rooms and a general diesel generator);
- for emergency cases spare rooms were prepared.

VI.3 Organisational security measures

22. Organisational security measures were as follows:

- information security management and policy, including drafting and implementing of legal acts;
- check of observance of security measures, for this purpose a population census security officer was appointed and his job description approved. The main duties and tasks of the security officer included checking of complying with the security requirements, looking through and analysing the security diaries, submitting regularly the security reports to the Director General;
- inspectors of the DPI were of great help in the check;
- staff-related measures, i.e. employing and training of computer operators (both in subject matter and technically), skill conversion, employing of shift managers, work field-based reporting and subordination, specification of access in job descriptions in accordance with the field of work, etc.;
- planning of action in emergency and reproduction cases, also in avoiding and informing of any trouble (order of informing, supply with relevant telephone numbers, etc.);
- implementation and observance of the bylaws of the Population Census Division;

23. The bylaws of the Population Census Division established:

- organisation of labour in the two-shift working regime;
- conditions for the access to and stay in the rooms;
- the obligation to register all the persons staying in data processing rooms, and the obligation by these persons to bear a nameplate;
- restriction to the access of visitors to the rooms and registration of visitors;
- registration of events in the diaries of the shift manager and managers;
- regulation of breaks;
- procedure for key management and observance of this procedure;
- appointment of a shift manager, a co-ordinator (appointing the shift manager contributed much to the functioning of the bylaws and security measures).

24. All these measures were applied to protect assets, to guarantee the handling, integrity and confidentiality of data, to ensure compliance with the requirements of the Personal Data Protection Act.

25. The Statistical Office allocated considerable funds to work out the Population Census security policy and apply security measures. For this purpose, contracts were concluded with the most highly acknowledged specialists of the field in Estonia. Data processing rooms underwent repairs, stern organisational, physical and infotechnological security measures were applied. Strict internal rules the observance of which was the task of shift managers and the Population Census security officer were established to be followed in data processing rooms.