

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

CONFIDENTIALITY IN STATISTICS LITHUANIA

Contributed paper

Submitted by Statistics Lithuania¹

1. Confidentiality has become one of our priorities in statistical policy of Lithuania. On 23 December 1999, the Seimas of the Republic of Lithuania passed the Law on Statistics, which described the definition of confidential statistical data, principles of statistical data dissemination and etc. On 17 July 1999, the Seimas of the Republic of Lithuania passed Personal Data Act, which regulates the protection of personal data.

2. Data security is imposed at the original and aggregated data levels as well as when releasing data in publications. To implement confidentiality requirements, instructions and measures, ensuring data security and confidentiality, have been worked out. The gist of the principles is that 3 levels of security and confidentiality have been developed: physical, legal and technological. In these levels there is the whole complex of means and documents, that ensures data security and statistical confidentiality:

3. Since 1999 a new job position has been created in Statistics Lithuania, namely, the position of the authorised envoy for data security and confidentiality. There are persons responsible for confidential data protection means in Statistics Lithuania approved by a special order in 1999.12.13, who prepare yearly data protection means plan and control the implementation. Also, there is a data protection service in Statistics Lithuania, which ensures data protection in physical level.

4. There are internal documents, which regulate data protection in our office:

Yearly confidential statistical data protection means plan;
Data protection on the networks;
Confidential statistical data protection order on physical level;
Permittance routine rules;
Rules for internal network.

5. All questionnaires enclose a note on data security and confidentiality guarantee. All employees, who work with confidential data, sign a deed of covenant. Data security and confidentiality principles enforcement in publications as well as those on the website is the object that is imposed on the Division of General Statistics. Specific regulations on statistical confidentiality are to be worked out, which will define principles of confidentiality more exactly. Confidentiality methods and means will be described there.

6. At the moment we are starting to use various confidentiality methods in Statistics Lithuania. The methods used are different for tabular data and for microdata.

Methods used for tabular data:

- Geographical thresholds – limiting geographical detail by releasing data only for areas above a particular spatial or population threshold.

¹ Prepared by Sigitas Biciunas (e-mail: Biciunas@mail.std.lt).

- Number 3 of respondents in survey – threshold rule where a cell in table is defined to be sensitive if the number of respondents is less than 3.
- If the threshold is not achieved we apply cell suppression.
- Random or controlled rounding – rounding to the top/bottom value using random rule or marginal rule.
- (n,k) rule - Regardless of the number of respondents in a cell, if a small number (n or fewer) of these respondents contribute a large percentage (k percent or more) of the total cell value, then the so-called n respondent, k percent rule of cell dominance defines this cell as sensitive. (only for data concerning enterprises). We defined such values - when n=1, then k=70%; when n=2, then k=85%.
- Recoding – recoding variables into broad categories to reduce detail. (only for data concerning enterprises)

Methods used for microdata:

- Geographical thresholds – limiting geographical detail by releasing data only for areas above a particular spatial or population threshold.
- Top and bottom coding – setting top-codes and/or bottom-codes on continuous variables. A top-code for a variable is an upper limit on all published values of that variable. Any value greater than this upper limit is not published on the microdata file. Similarly, a bottom-code is a lower limit on all published values for a variable.
- Sampling – releasing only a small proportion of the original data as a microdata file.
- Recoding – recoding variables into broad categories to reduce detail.
- Deletion of especially sensitive records – in microdata files, certain statistical units may be particularly sensitive. These may be excluded or deleted from the microdata file to be released.
- Deletion of data items – certain hard to protect or particularly sensitive information may be deleted from particular microdata records.
- Blur the data – aggregating across small groups of respondents and replacing one individual's reported value with the average.
- Micro aggregation - records are grouped based on a proximity measure of all variables of interest, and the same groups of records are used in calculating aggregates for those variables.

7. All these confidentiality methods are not used effectively because of the lack of software. We don't have any special software to limit the disclosure of confidential data in relation to the release. So all these methods are used manually. It would be very useful to obtain training in such software, to compare the differences and to find out the possibilities of such programmes.

8. Another problem that we have encountered is the small country problem. There are just 1-2 monopolistic enterprises in some economic activities. Keeping to the main principles of confidentiality is indispensable. But if we don't publish any information on these economic activities, statistics will not meet its goals. We don't know how to coordinate confidentiality requirements and our needs. It would be very useful for us to get consultation on these problems. We look forward to receiving any recommendations or offers of a way out of this situation at the meeting.

9. Participation in EU activities related to data security and confidentiality, study visits to explore progress in the development and implementation of the data confidentiality methods and techniques in the statistical agencies of transition countries and small EU member countries would indeed be requested.