

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

**PROGRESS IN THE IMPLEMENTATION OF SDC METHODS AND TECHNIQUES IN
CENTRAL AND EASTERN EUROPE**

List of key issues for discussion

Submitted by Statistics Netherlands¹

1. In this topic the progress is studied in the development of data confidentiality methods and techniques in the statistical agencies of the transition countries. The topic has attracted four invited and nine contributed papers. Invited papers were submitted by Czech Republic (30), Estonia (31), the Former Yugoslav Republic Of Macedonia (32) and the UN ECE Secretariat (43). Supporting papers were submitted by the Former Yugoslav Republic Of Macedonia (18), Armenia (33), Azerbaijan (34), Hungary (35), Kazakhstan (36), Kyrgyzstan (37), Lithuania (38), Mongolia (39) and the Ukraine (40). The papers contain both positive and negative news on the implementation of SDC methods and techniques. Large differences exist within Europe in the progress of the implementation of SDC methods. Both successful experiences and problems faced when developing projects on data confidentiality on the national level are described in the papers. It is quite clear that different countries can learn a lot from each other. Theoretical knowledge on SDC methods and experiences on the implementation in the production of official statistics should be shared. Reading papers of other European countries is a useful approach. Listening to each others presentations and actively participating in the discussion is another stimulus. Requests for international assistance in this field will also be explored in the general discussion.

2. Before one can implement SDC measures successfully one must have the theoretical knowledge. Many statistical offices have some people who know enough about SDC to start the implementation phase. However, it would be good to spread the knowledge more widely first. It is not very attractive for a statistical office to become dependent on one or a few persons. By training the general level of knowledge on SDC could be extended. The international training courses on SDC give excellent opportunities to both refresh existing knowledge and spread the knowledge over more people working for statistical offices. Also national courses, seminars and lectures on SDC and last but not least the Joint ECE / Eurostat Work Sessions on Statistical Data Confidentiality help in reaching this aim. As soon as the level of the knowledge of SDC has reached a reasonable level a statistical office can start the practical implementation of SDC measures. It is not wise to try to change everything at once. The most successful approach is certainly improving the data confidentiality practice by several successive projects.

3. A necessary requisite for a successful policy on SDC is a law on statistics. Many transition countries have had changes in their laws on statistics in the past ten years. Hungary (35) writes that their current regulations have been strongly influenced by the transformation process of 1989. Gradually, Hungary has reached the EU standards on protection of personal data. Ideally, as in Armenia (33) the law on statistics regulates the statistical confidentiality by introducing the framework for the storing, using and disseminating of statistical data. Regulations, procedures and software measures complete the total picture of the statistical confidentiality. It is important to discern the data streams from respondents to the statistical office, between different branches of a statistical office and from the statistical office to the

¹ Prepared by Eric Schulte Nordholt (e-mail: ESLE@CBS.NL).

users as Azerbaijan (34) does. In the Ukraine (40) the statistical legislation and legislation on data protection has improved last ten years. The respondent in the Ukraine has the right to check and if necessary correct his or her own record in the census.

4. After a number of years of research in SDC measures we gradually have moved to sets of standards of confidentiality of microdata. Lithuania (38) describes many different measures they use to protect the microdata. They reduce the level of detail in the identifying regional variables, apply top- and bottom-coding, use sampling as a protection measure, recode some variables, delete sensitive records, blur the data (e.g. replace a value by its group mean) and finally make use of micro aggregation techniques. Not all countries will use all these techniques, but it is important to be aware of all the possibilities that exist to protect microdata so that as much detail as possible can be released without disclosing individual sensitive information.

5. Also in the field of the confidentiality of tabular data we have reached the point that a set of standards is available. Lithuania (38) gives an overview how tabular data can be protected. Identifying regional variables can be recoded so that less detail is given but the data will be better protected. A cell is considered sensitive if it has at most three contributors. Of course one can also choose another threshold value, but this does not change the principle of defining some cells as sensitive. After having identified which cells are sensitive, these cells are suppressed. For enterprise data the dominance rule is applied. This rule states that a cell is regarded sensitive if the sum of the largest n contributors account for more than p percent of the total cell value. Finally, Lithuania claims also to use top- and bottom-coding and recoding techniques for enterprise data.

6. Many countries have been active in recent years in improving the physical protection of their computer networks. The chain of statistical protection has a weakest link and it is wise to start with an adequate physical protection of the sensitive data before the statistical protection is improved. Kazakhstan (36) discusses the additional protective measures of testing regularly firewalls, organising the storage of backup copies and limiting password validity. The Former Yugoslav Republic Of Macedonia (18) describes in more detail how one can improve the rules for creating a password, how one can define the schedule for password changes and what kind of users there are and which privileges they have. Physical and statistical protection is not free of cost. Therefore Azerbaijan (34) states rightly that a reasonable level of protection is wanted given the financial constraints one has to meet.

7. In Hungary (35) the departments publishing the data are responsible for the correct application of SDC measures. They have organised the process in such a way that advices can be obtained from legal, technical and statistical advisers. Kyrgyzstan (37) claims the importance of the independent status of their statistical office. There is an important link between the independent status of a statistical office and the policy on SDC. Only an independent statistical office can namely thoroughly protect confidential information. In Mongolia (39) a revised law on statistics established the independent status of the statistical organisation and increased administrative sanctions in cases of breach of the law.

8. A major problem for most transition countries is the availability of software for statistical disclosure control. Hungary (35) started the testing of the ARGUS software after the Joint ECE / Eurostat Work Session in Thessaloniki in 1999. The ARGUS packages are free available, so there is no financial problem in testing and applying them to protect microdata and tabular data. However, some training in the use of these packages in the practice of official statistics is wanted by many countries. Lithuania (38) writes that all the confidentiality methods are not used effectively because of a lack of software. Therefore, they currently use all methods manually. As a small country they have an additional dominance problem: often just one or two monopolistic enterprises dominate in some economic activities. Suppression of such a cell would imply that the statistical tables published would not meet its goals. Maybe individual consents of some of these enterprises could solve this problem.

9. Some countries need more theoretical knowledge about statistical disclosure control. Other countries have enough knowledge but lack software to maintain protection rules effectively. Moreover, it takes time to test new software and become an efficient user of e.g. the ARGUS packages. Also for this purpose training is needed. The need of training in SDC is thus clear for most countries, although the kind of training may vary among countries. The Ukraine (40) would find it extremely useful to learn about the development work and results of the implementation of data protection systems in the statistical agencies

of other countries. Lithuania (38) states that they would like to participate in EU activities related to data security and confidentiality. They request 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.

10. An important part and often neglected part of SDC in practice is the communication with the citizens. They want more and more information from the statistical offices. Therefore, it is important that they participate in the surveys and give the correct answers. Only if they feel that their data are protected carefully they want to continue to participate in surveys. Hungary (35) initiated a special communication program related to this topic with special respect to the 2001 Population Census. Even in case participation is obligatory the statistical office remains responsible for the protection of the privacy of the citizens and enterprises in the country. If people get the feeling that the protection has not been organised adequately, we face the risk that they are less careful in the answers they give about themselves and the enterprises where they work. Kyrgyzstan (37) writes about a well-known problem that is faced in most countries. There is a considerable growth in the demand for statistical data and at the same time a decrease in the willingness to provide information. Citizens complain about the administrative burden and statistical offices face the non-response problem. The duality of the situation makes the citizen less trusting towards statistics. We have to be aware that there is a risk that many citizens show scepticism about data confidentiality. Therefore, we have to take SDC seriously and should not get tired of explaining what is protected and why this is the case.