

Introduction

Protecting the confidentiality of statistical data is one of the fundamental principles of official statistics. Data protection issues have always been at the top of the agenda of statistical agencies and undoubtedly have contributed to the inception of the “Fundamental Principles of Official Statistics” in 1992 of the Conference of European Statisticians.¹ With the expansion of Internet, a growing amount of data available from various sources (including administrative registers and government-wide Extranets), and an increasing demand for more detailed data, data protection issues are becoming even more important.

In practice, statistical offices have to establish a satisfactory balance between the data users’ need for detailed data and the respondents’ right to privacy and confidentiality. On the one hand, statistical agencies devote considerable resources to data collection and are interested in making these available to as wide an audience as possible. Pressure to provide more data is coming from researchers as well as policy makers. On the other hand, protecting the privacy and confidentiality of respondents is indispensable in gaining cooperation in surveys. A breach of confidentiality will cause the mistrust of respondents, which can result in decreasing response rates and deterioration in the quality of collected data. Another important aspect is the public perception of confidentiality. The legal, organisational and technical protection measures will only enable statistical offices to obtain high-quality responses if the public both knows about and believes in data protection methods.

There is a long route from the general principles of statistical data confidentiality (SDC) to their implementation, starting from the legal and institutional prerequisites (such as Laws on Statistics, data protection, etc.) to the practical methods and software. Even when rules and tools are in place, changing circumstances often make it necessary to review both regularly to ensure their effectiveness and to assess the impact they have on data quality. The current issue of the Statistical Journal aims to shed light on some of the latest developments in this area.

Confidentiality problems have been discussed at the international level for over a decade, and several international scientific conferences have been dedicated to that topic. Since 1999, UNECE and Eurostat have jointly organised biannual Work Sessions on statistical data confidentiality with the emphasis on the application of disclosure control in statistical practice. As the membership of UNECE includes countries which are well advanced in SDC and countries just beginning to deal with these issues, it offers a forum to exchange experiences and to learn from the best practices of the most advanced countries.

The current issue of the “Statistical Journal” contains a selection of key papers presented at the Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality held in Skopje, The former Yugoslav Republic of Macedonia in March 2001. No less than 37 countries, as well as many international organisations and universities were represented and 50 papers were submitted to the meeting.

The wide range of articles presented in this issue span from general SDC policies to research in specific technical methods. The first set of papers deal with the confidentiality policies of a statistical office. The paper by Bo Sundgren serves as a good introduction to confidentiality protection in statistical agencies which attempts to “strike a balance between the protection of privacy of the respondents and freedom of

¹<http://www.unece.org/stats/archive/docs.fp.e.htm>.

information". It discusses the considerations behind an office's confidentiality policy and recommends some directions for solving the problems. Many different legal, methodological and administrative tools are available but it is best to combine them in such a way that will allow increasing the use of detailed data and confidentiality protection at the same time.

The paper by Gerald Gates and a joint paper by Nick Greenia, Julia Lane and Diane Willimack examine the attitudes of respondents and the general public toward confidentiality. The ability of statistical agencies to reassure the public that their promises of confidentiality have been kept is sometimes as important as the disclosure control technology. The use of technical measures is not sufficient in itself when the public and respondents do not understand nor believe in these measures. The public perceptions of confidentiality, especially those of businesses, are a comparatively less studied area in SDC. The knowledge of what data are perceived as confidential by respondents and how the passing of time affects data sensitivity helps to find the most efficient way to protect the data. If some data are less sensitive than others, resources can be focused on those data elements that are more sensitive for respondents.

Anco Hundepool gives an overview of the development of Argus software under the project on Computational Aspects of Statistical Confidentiality. Argus with its two components: μ -Argus for the protection of microdata and μ -Argus for tabular data is a widely applied disclosure control software that is available for all statistical offices free-of-charge. Hence, it is useful to know what developments in the software can be expected in the future.

The two papers, by Eric Schulte Nordholt and by Marja Tammilehto-Luode, describe the guidelines used for the implementation of disclosure control in statistical offices. These two papers are good examples of what kind of general disclosure control rules are needed and how the software packages are implemented.

Another set of papers deals with research in specific computational methods of disclosure limitation. New methods of confidentiality protection or new uses of traditional methods are explored, such as: model-based disclosure limitation using imputation (by Luisa Franconi et al.); microaggregation (by Anna Oganian and Josep Domingo-Ferrer); cell suppression (by Matteo Fischetti and Juan-José Salazar-Gonzalez); and releasing bounds for cell entries in the case of series of statistical queries on a database (by Adrian Dobra and Stephen Fienberg).

The following two papers consider risk assessment: re-identification risk by matching register and survey data (Johann Bacher et al.), and by simulation of intruder behaviour (Mark Elliott). Risk assessment plays an important role in deciding how the confidentiality protection measures should be applied on different datasets. By estimating the disclosure risk, it is possible to choose methods that increase both the detail of the released data and the level of confidentiality protection.

Finally, the last paper (by Tiina Luige) gives an overview of the developments in the area of statistical data confidentiality in the transition countries based on a survey conducted by the ECE secretariat in 2000/2001. Although SDC issues have only been dealt with since the beginning of the 1990s, the legal basis for confidentiality is in place in most of these countries. The statistical offices recognise the importance of confidentiality protection and have made good progress in implementing the organisational and administrative safeguards for disclosure control. The use of mathematical methods is less advanced because of lack of knowledge of appropriate methods. International conferences, such as the one held in Skopje, offer opportunities to overcome this disadvantage.

It is important to state that there were many interesting and high-quality papers presented at the meeting which, unfortunately, could not all be included in this volume. Interested readers can therefore consult the meeting's Web page where all documents are available at: <http://www.unece.org/stats/documents/2001.03.confidentiality.htm>.

Tiina Luige
UN Economic Commission for Europe
Geneva
Switzerland