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

THE APPROACH TO STATISTICAL CONFIDENTIALITY IN THE NATIONAL STATISTICAL SERVICE (NSS) OF THE REPUBLIC OF ARMENIA

Contributed paper

Submitted by the State Statistical Council of the Republic of Armenia [1]

I.   INTRODUCTION

1.   Statistical data must reflect as accurately as possible the socio-economic situation in society in order to meet the needs of users of statistical information.

2.   If quality statistical data are to be obtained, it is essential to have the trust of respondents. This trust must be based on legal aspects whose purpose is to protect respondents and prevent the information provided to the statistical service from being compromised.

3.   The legal basis for establishing relationships based on trust between respondents and the NSS of the Republic of Armenia is provided by the Law of the Republic of Armenia on State Statistics, adopted on 4 April, 2000, and the Law of the Republic of Armenia on the Population Census, adopted on 12 October 1999. The annex to this document contains articles from these laws concerning the maintenance of statistical confidentiality when working with statistical information and the confidentiality of information provided by respondents.

4.   These laws pay a great deal of attention to the confidentiality of primary data and provide for the legal liability of statistical service workers if they do not observe the procedure for storing, using and disseminating statistical data subject to statistical confidentiality.

5.   To ensure that the rights to statistical confidentiality guaranteed by the legislation are observed, they have to be backed up by regulations, procedures and software measures.

6.   A decision of the State Statistical Council is being prepared regarding the procedure for using, storing and disseminating confidential data. Once this document has been adopted it will be published as a departmental regulation. It will detail the standards of confidentiality set out in the Law on State Statistics. It will define two types of data: individual data (microdata) and aggregate data (tabular data), for which there are separate definitions of confidentiality. Microdata are confidential when direct or indirect identification of individual (personal) information is possible, whereas tabular data are considered confidential if two individual components account for more than 75% of the total.

7.   The Regulation also provides for designating the person responsible for storing primary data. There is also provision for setting up a committee to monitor confidentiality issues.

8.   The State Statistical Council adopted the Decision of 15 August 2000 on the Basic Responsibilities of NSS Staff, section 5 of which "Responsibilities of the Service"(NSS of the Republic

---

[1] Prepared by V.G. Davtyan and A. Shaboyan.

of Armenia)" refers to maintaining statistical confidentiality (5.1) and providing a mechanism for the non-dissemination of statistical data without the agreement of the person who provided the data (5.2).

II.   CURRENT SITUATION

9.   Issues of confidentiality in the NSS of the Republic of Armenia are currently governed by the methodology developed in the Methodology department on the basis of the above-mentioned laws and regulations. The NSS of the Republic of Armenia is at the stage of drawing up and adopting regulations defining the procedures for NSS staff for working with information subject to statistical confidentiality, particularly information on electronic media.

10.   In our view, the measures required should include the following:
* Developing a system which protects information in the NSS and sets out the procedure for NSS staff for working with information subject to statistical confidentiality.
* Developing software measures for defining the procedure for working with statistical data and information on electronic media, and the procedure for storing this information in electronic form and for its dissemination.
* Training administrators (setting up an institute of administrators of electronic information resources) on SQL, Novell and WEB servers.

11.   On entering the service, all NSS staff must sign an undertaking regarding statistical confidentiality applying over a number of years and after they have left the service.

12.   If the needs of users of statistical information are to be met, there has to be a suitably developed infrastructure in place to allow the exchange of a huge volume of data both between the NSS offices located throughout the Republic of Armenia and between the NSS and data users. To achieve this, the procedure for entering, correcting, processing, transferring and disseminating information has to be automated, whereby great attention has to be paid to the issues of data protection, since when it is being stored, transferred and processed in electronic form, or if there is careless handling or inadequate protection provided, the information could easily be open to unauthorised copying and dissemination at any stage. In order to implement the measures on data protection, it is therefore necessary to consider in detail the whole cycle from input to dissemination of statistical data.

13.   Primary information (microdata) are entered by operators using input programs based on MS Access, MS Excel, Visual Basic 6.0, and the information is then subject to a logic check and corrected. The databases are managed in the corresponding departments, the heads of which are responsible for the procedure for working with them. Copies of the databases containing microdata are located on the Novell server, access to which is strictly regulated. The data are subsequently modified to take account of confidentiality requirements but without compromising the principle of the completeness of statistical information. The data are then located on the SQL server, the administrator of which establishes, in accordance with the regulations, the procedure for, and level of access to, these data for individuals and takes measures regarding:
* accessibility, i.e. obtaining the required information within a reasonable time;
* completeness, i.e. using the available software to protect the information from unauthorised changes and destruction, and ensuring that it is up to date (procedure for updating information) and that there are no discrepancies;
* Confidentiality, i.e. protecting the data from unauthorised consultation, and the timely detection of unauthorised consultation.

14.   We would also point out that information is also protected from destruction by the procedure for archiving and restoration from archived copies. Here the regulations have to define the periodicity of the archiving of the information and the procedure and length of time for storing archived copies.

15.   For each user on the server the administrator establishes a profile which defines the level of access to particular information. Access is defined separately for each directory, and only one user is entitled to change the information in a given directory: for the other users (if there is open access to the information according to the regulations) access is on a read-only basis. Moreover, MS Excel establishes individual passwords for each of the local databases, which are generally based on MS Access.

16.   There may be one password to access a particular database for consulting and working with it, and another password for entering changes.

17.   Once the microdata have been processed using the algorithm developed for each individual case, and the aggregate data (macrodata) and output tables have been obtained, the data have to be published and disseminated in compilations and analytical reports and on the Internet etc.

18.   It should be borne in mind that the most advanced software may be used to identify statistical accounting units from the published aggregate data.

19.   In order to exclude the possibility of direct or indirect identification of persons providing statistical information, a special algorithm has to be used to check the data prior to publication in order to be satisfied that the output information does not directly or indirectly contain information subject to statistical confidentiality (i.e. that the output information meets the confidentiality criteria). In order to do this, a checking algorithm has to be developed for each output table. The Methodology department intends to develop the checking algorithms jointly with the Aggregate Data department, and appropriate software has to be developed for this.

20.   At present we have started developing an inquiry system for the databases on the SQL server, access to which will be through the www.armstat.am site. The algorithm will be designed to provide inquiries which include confidentiality checks of output tables.

III.   CONCLUSION

21.   Since the consequences of compromising the principle of confidentiality may be extremely serious and may undermine respondents' trust in the Statistical Service for many years, it is necessary right from the start to pay great attention to the confidentiality of statistical data.

22.   As can be seen from the legislation, data confidentiality is guaranteed for practically everyone, except those who operate a monopoly.

Annex

I.   Extract from Article 14 of the Law of the Republic of Armenia on State Statistics:
"Article 14. The collection, use, publication and storing of state statistical data and other information.

... Statistical data shall be considered confidential (subject to statistical confidentiality), if they can be used directly or indirectly to reveal (determine or identify) a provider of statistical information and obtain individual (personal) information. Data which are publicly available from other sources shall not be considered subject to statistical confidentiality.

Data which are considered subject to statistical confidentiality shall not be published or in any other form disseminated without the agreement of the information provider. Agreement must be given in writing. In the case of legal persons, agreement shall be given by its executive body or other duly authorised body. The agreement may contain restrictions on how the data are used. The provision for not publishing or in any other form disseminating information shall not apply in the case of information relating to natural or other monopoly activity, the non-publication (non-dissemination) of which would materially affect the maintenance of the principles of the completeness of statistical information. In such cases the Council shall decide on publication (dissemination).

Direct identification of information providers takes place when the provider's name (designation), location and classification identifying code becomes known. Indirect identification takes place when it becomes possible to discern the individuality of a particular unit using data which are not listed above.

The Council may decide that, for scientific research, data subject to statistical confidentiality may in exceptional cases be used or published in part provided that individual (personal) data cannot be directly identified.

The reporting or providing of collected information to staff members of the statistical body and persons serving this body shall not be considered as publication or dissemination if:

a) these data are required for carrying out statistical work;

b) these persons undertake not to publish the confidential information entrusted to them or made known to them in connection with the work, and undertake not to use or publish the data for their own benefit or that of third parties.

In accordance with the procedure established by the Council, the statistical body shall take measures for the appropriate storing of information subject to statistical confidentiality.

The Council shall determine the forms in which statistical information is to be published, in accordance with the procedure established by laws and other legal acts.

... The procedure and time periods for storing state statistical information shall be determined by law."

II.   Extracts from the Law of the Republic of Armenia on the Population Census, Article 12
"The use, publication and storing of information obtained in the course of the population census"

"The final results of the population census shall be used or published only in aggregate and grouped form.

Primary (individual) data from the population census shall remain confidential and shall not be published."

Article 13. "Legal liability for violation of the legislation on the population census" reads as follows:
"Staff coming into possession of individual information on the population are forbidden to communicate to anyone the content of the completed population census documents. In the event of an infringement of the procedure for using primary (individual) population census data, population census workers shall be held liable under the procedure established by the legislation of the Republic of Armenia. Judicial, legal investigation and other bodies may not use the primary census material as evidence."