

Topic (iii) - Administration and policy of statistical data confidentiality

**ADMINISTRATION AND POLICY OF STATISTICAL DATA CONFIDENTIALITY IN
ISRAEL**

Submitted by the Central Bureau of Statistics of Israel¹

Contributed paper

I. INTRODUCTION

1. A primary duty of the Israeli Central Bureau of Statistics (ICBS) is to collect and process statistical data and to disseminate the results. To carry out these duties, the ICBS has the legal authority to collect statistical data from both public and government institutions. Furthermore, according to the Law on the Protection of Privacy of 1981, the ICBS has a duty to protect confidentiality of the collected individual data as set out in the Statistics Ordinance. In addition to the statutory duty, the ICBS has its own interest in maintaining the data confidentiality, since a breach of confidentiality could jeopardize the public cooperation necessary for carrying out ICBS's functions.

2. The dissemination of the results of ICBS's statistical activities encompasses a wide variety of statistical information products made available to the public. These products can be divided into two categories: aggregated information (mostly on paper), and microdata files. As computing power has increased, so has the demand for more detailed microdata files, especially from the academic community. In order to meet this demand, the ICBS has developed procedures which make possible both the release of microdata files and protecting data confidentiality. This paper provides an overview of the procedures for ensuring confidentiality of the statistical data released by ICBS. The paper focuses on the safe release of microdata files and not on computer (and information) security which, although related to the subject of confidentiality, is in a category of its own.

II. LEGAL BACKGROUND

3. Confidentiality of individual data collected for statistical purposes has been protected in Israel since 1947 in accordance with the Statistics Ordinance, under which the ICBS has operated since its beginning. On the subject of confidentiality, the Ordinance (which was last amended in August 1978) states:

"No information collected for the purposes of this Ordinance and derived from an individual reply or an answer to a question, or from records or documents as referred to in section 15, shall be so published as to enable the identification of the person to whom it relates". [Art. 17(b)]

¹ Prepared by Gideon Burshtein.

And:

“No person other than an employee [of the ICBS] shall see any individual reply made for the purposes of this Ordinance, or any part of such a reply, except for the purposes of a prosecution under this Ordinance”. [Art. 17(c)]

III. HISTORICAL BACKGROUND

4. Until the 1970s, maintaining the confidentiality of individual data was not a serious problem. Microdata files were almost never released. In addition, the computers available to researchers in those days had limited computing power. As for aggregate data, usually published on paper, the ICBS used some simple but efficient “rules of thumb.” For example, no statistical table could be disseminated unless there were at least 3 cases (sample observations) per a table’s cell. Thus confidentiality was secured, though sometimes at the price of data suppression.

5. In the 1970s, and more so in the 1980s, the ICBS began releasing microdata files containing anonymous data in response to an increased demand, especially from the scientific research community. During that time there was still little difficulty in supervising the release of such microdata, since all of the ICBS computerized data files were stored in one central mainframe computer. The Government Statistician, who directs the ICBS, was the only person with authority to approve the infrequent release of such files. Thus, supervision of released files was centralized both technically and administratively.

6. Microdata files were released without any information that would enable **direct identification** of an individual (i.e., a name or an ID. Number). Responses that might enable an **indirect identification** (such as detailed occupation or the year of birth) were grouped into broad categories to prevent disclosure. These techniques were sufficient to maintain confidentiality under the still-limited technological capabilities available to researchers at that time.

7. Today, as the result of the great advances in information technology, the situation has changed. The computers of today, including personal computers, have placed in the hands of researchers tremendous computing power unavailable 15-20 years ago. Another development has been the change in the organization of work at the ICBS itself. There has been a great increase in the use of PC’s in the ICBS, with the inevitable result that many microdata files are stored on staff members’ individual PC’s. This development has been accompanied by some increase of decentralization in working procedures, as well as in the potential capabilities of dissemination of files.

8. These changes have given rise in the ICBS to the necessity for dealing with two aspects of ensuring confidentiality. The first one is that of “computer security” per se, which, of course, has in its turn an influence on confidentiality. The second one is that of establishing a clear and concise policy – and a suitable operational administration – for release of microdata files from a decentralized computing environment while ensuring confidentiality.

9. Computer security has been obtained in the ICBS by using technological means that are out of scope of this paper. In order to formulate a clear policy, the Government Statistician appointed a special Statistical Confidentiality Committee (SCC), headed by the ICBS Chief Scientist. The function of the SCC has been to draft a concise **confidentiality policy** for dissemination of information, and then to handle its implementation. Release of microdata files was the urgent problem, but data in these files was not the only kind of information to be monitored.

IV. ICBS POLICY FOR RELEASING MICRODATA FILES

10. The SCC began its work by formulating policy guidelines for safe release of microdata files. It presented its recommendations to the Government Statistician in 1995. After these guidelines were approved by the Directorate of the ICBS, the SCC began their implementation on a day to day basis.

11. The underlying principle of the ICBS policy is that statistical data is collected and produced in order to be used. In fact, “statistical data dissemination” is one of the duties of the ICBS under the Statistics Ordinance [Art. 3(1)], as long as it does not involve a breach of confidentiality.

12. It was clear that it would be preferable to find a way to identify any risk of disclosure and thus to enable release of all other “safe” data. The SCC was unable to propose a secure and inexpensive method to pinpoint the exact probabilities that particular individual respondents could be identified. Solving this problem appeared to be too complicated and expensive. Therefore, the committee had to be content with formulating a working standard, by which keeping statistical confidentiality means ensuring that the possibility of identifying any individual is highly improbable, though not totally impossible. The SCC found the alternative, of requiring that identification by active search must be impossible, to be an unrealistic goal in this age of very advanced computers. It was clear that if the ICBS tried to implement such a requirement, it would prevent the release of any microdata files to the public.

V. GUIDELINES FOR SCC POLICY IMPLEMENTATION

13. Examining ways to minimize the chances of disclosure in microdata files, the SCC proposed several solutions. It recommended that practical prevention rules, based mainly on intuitive considerations and common sense, would be sufficient. Therefore SCC has established the following guidelines for the release of microdata files, which are meant to ensure compliance with the above policy. The following rules are intended to ensure that the microdata file, in and of itself, is “immune” from a reasonable possibility of disclosure.

- (i) **No direct identifying data** – a released microdata file must not contain any direct identifying variables (such as name, address, ID number). While this is obvious, it does need to be stated.
- (ii) **Limiting of variables and categories** – when building a file for a specific research project, only relevant variables should be included in the microdata file. One can never tell when superfluous data might enable identification.
- (iii) **Grouping variables** – grouping of very detailed responses (to some variables) into broad categories, while somewhat limiting the value of the file, is appropriate especially for those microdata files made widely available to the public.
- (iv) **Splitting files** – another recommendation is to divide information released from censuses, or from big surveys, into more than one file. For example, a file might be released containing detailed demographic data, but with very limited geographic data, or vice versa.
- (v) **Combining any of the above means with contractual limitations** – the “immunity” of files might also be achieved by imposing contractual restrictions on the release and handling of microdata files that in and of themselves are not quite “immune,” in order to ensure the requisite level of statistical confidentiality.

14. The following means of ensuring statistical confidentiality in microdata files were examined by the SCC, and rejected.

- a) **Noise** – the addition of noise to microdata files, reducing the possibility of disclosure by changing the data, was found unacceptable by a special committee of users appointed by the Public Advisory Council for Statistics (a body that operates under the Statistics Ordinance).

- b) **Special data processing service** – the establishment of a special fast and efficient data processing service, provided by the ICBS for researchers, would have enabled use of detailed microdata files without compromising statistical confidentiality. This solution was rejected since researchers prefer to “get a feel for” the data by carrying out exploratory analyses that depend on a rapid turn-around time. Therefore, this way would never be sufficiently fast or convenient for the research.

15. The SCC also stipulated that in principle, **no** data about business or organization units (not even unidentified data) could be released, unless the explicit consent of the respondents was obtained. Therefore, research utilizing such information has been practically limited to the Research Room, mentioned below.

VI. FORMS OF RELEASE OF MICRODATA FILES

16. Considering the means selected for minimizing disclosure chances, the SCC established three ways to enable use of microdata files by researchers. The first two ways deal with dissemination of files, and the third one with access to files on the premises of the ICBS. Today the ICBS is releasing microdata in the following forms:

- **Public Use Files (PUF)** – these are microdata files containing anonymous (unidentified) individual records, with responses sometimes grouped into broad categories, to prevent disclosure. These files, containing relatively little detail for individual variables, meet the standard of making disclosure highly improbable, and thus can be released to the public.
- **Microdata Under Contract (MUC)** – anonymous microdata files containing individual data that might be more detailed than that of PUF. By its structure a MUC file might not comply with the above standard; but it still prevents disclosure because of additional limitations imposed on its release and use. Release of MUC files is restricted to academic researchers in recognized (listed) institutions, and it is done under contractual limitations governing the use of the file by these researchers. There are some 10 terms in the agreement that deal with maintaining confidentiality, such as forbidding use of the file by a third party or banning any attempt to match the file with another microdata set.
- **ICBS Research Room (RR)** – a special site within the premises of the ICBS, in which researchers from recognized academic institutions are able to carry out their work getting access to anonymous but otherwise uncensored microdata files. The use of the RR is the only way researchers can get access to data of business or organization units. Confidentiality is achieved by a combination of physical, administrative and legal terms:

17. Requests from researchers from recognized institutions who wish to use the RR facility, must be screened. The research has first to be endorsed by the ICBS Chief Scientist. After the research is approved, the researcher has to become a special sworn employee of the ICBS. He is thereby subject in his work on ICBS data to the Statistics Ordinance and its criminal sanctions for breach of its confidentiality requirements. The ICBS imposes strict controls on the ability of researchers to remove from the RR output in any form.

VII. ADMINISTRATION OF MICRODATA RELEASE

18. The initial release of every microdata file to a researcher, either as PUF or MUC, must be approved by the SCC. All PUF files are released subject to a license of use. All MUC files are released only after a strict Researcher’s Agreement is signed. The use of a microdata file in the Research Room must be approved by the ICBS Chief Scientist. Upon his approval for the research, the researcher must sign an agreement with the ICBS ensuring compliance with the rules and regulations applying to the use of the Research Room. Only then does the researcher sign the documents making him a special sworn employee of the ICBS and gain access to the microdata file in the Research Room.

19. At the **operational level**, the copying a microdata file from a PC to a floppy disk, for delivery to a researcher, requires a special conversion process. This process involves the use of dedicated computers and programs, at a special "station." The conversion process may only be carried out by authorized personnel, who will perform the conversion only upon receiving the prior written authorization of the ICBS Legal Advisor. This whole process ensures effective supervision over the release of microdata to the public and compliance with the ICBS confidentiality policy.

VIII. FUTURE PLANS

20. The next project on the SCC's agenda is to implement the advanced tools for assessing the probability of identifying individual information in a microdata file. Today, as mentioned above, assessment of the likelihood of disclosure is based primarily on personal professional judgment, with meager assistance from simple computerized checking. We hope to adopt or develop sophisticated tools and procedures that will give us more accurate evaluations of chances for disclosure of individual data in those microdata sets intended for release to public use.