



Economic Commission for Europe**Conference of European Statisticians****Sixty-first plenary session**

Geneva, 10-12 June 2013

Item 4 (b) of the provisional agenda

How should national statistical offices respond - moving from risk avoidance to risk management?**Improvements on risk management for micro-data access at the Brazilian Institute of Geography and Statistics****Note by the Brazilian Institute of Geography and Statistics****Summary*

The aim of this paper is to describe recent improvements on the procedures that have been implemented by the Brazilian Institute of Geography and Statistics, in order to minimize disclosure risk, to optimize the process to request online access for scientific use, as well as to give more transparency to the public, providing information about the methods available to access data, while still protecting data confidentiality of surveys and census.

The outstanding development of Information Technology in the last years has strongly changed the way that national statistical offices deal with the data users. On the one hand, national statistical offices are willing to meet different users' needs for data access, especially those coming from researchers, which require very detailed information and databases for their work. On the other hand, the national statistical offices have the legal and moral obligation to respect the privacy of the respondents. As a consequence, several national statistical offices have developed procedures and invested in solutions in order to find the right balance to face that challenge.

Following this trend, since 2003, the Brazilian Institute of Geography and Statistics has provided restricted access for researchers to data files that the statistical agency does not release to the general public, enabling a more in-depth analysis via online access at the headquarters of the agency.

* This document was submitted late due to delayed inputs from other sources.

I. Introduction

1. Statistical dissemination in the Brazilian Institute of Geography and Statistics (IBGE) was traditionally performed in two ways: for the general public, by means of media communication, press release results or conferences; and for the general users, through printed publications and electronic publications. Sometimes, when there are other needs of data coming from more specialized users and from government agencies, the requests are usually met through customized tables and public use micro-data files.
2. A policy of free dissemination of all products through Internet was adopted in IBGE, since 2001, which generated an outstanding increase in this communication channel. The number of pages of all paper publications has been decreasing since then, as the publications have an enclosed CD-ROM, presenting many additional tables to those in the printed volume.
3. Besides the electronic publications the IBGE web site offers two important databases: the Aggregated Database (SIDRA) - Database with information grouped in territorial level that allows the users to construct tables according to selected information - and the Multidimensional Statistical Database (BME) – Database with micro-data that allows users to construct tables according to selected information and confidentiality constraints. This database requires an internet subscription.
4. Furthermore, IBGE has been releasing public use micro-data files for households' statistics since early 1990s. Measures taken to protect the confidentiality of these micro-data include data suppression of geographical detail. On the other hand, there is no release of public use micro-data files for business data, neither for the Agricultural Census, nor for the short form of the Population Census data, which is used to collect information on the whole population.
5. Despite the wide dissemination of IBGE data through publications, web releases, etc., the inability to release files with more detailed data limits the usefulness of these data for research, policy, and programmatic purposes.
6. The increasing demand for highly detailed micro-data, the advance of technology and the increase of sensitivity to privacy issues have led IBGE, since 2003, to provide researchers with online access to data files which are not release to the general public, enabling more-in-depth analysis than the tabular aggregated data. In response to the public research community's interest in restricted data, IBGE established a Data Laboratory, a place where researchers can access detailed data files in a secure environment, without jeopardizing the confidentiality of respondents.
7. In these last 10 years, about 200 research proposals have been submitted and analyzed by the Committee of Assessment of Restricted Data Access. Of these 200, 73% refer to business surveys, most of which related to the Technology Innovation Survey. Other requests are related to Agricultural Census, Population Census and Household Surveys.
8. Recently, IBGE has recognized the need to review some guidelines, rules and procedures that have been adopted for researchers to gain online access to data files. Therefore, a broad study of the existing access procedures and practices developed by other national statistical offices has been undertaken, serving as a basis for the improvements that are to be implemented in the second half of 2013. This extensive review of best practices brought about changes in the procedures for Data Laboratory access as described below.

II. Current status and improvements

9. The current steps required to gain access to micro-data files for scientific purposes are described below, along with the changes and improvements that are to be implemented to the process.

A. Submission of research proposal

10. To access restricted data through the Data Laboratory, researchers must first submit a proposal. Currently, researchers submit their proposals without a specific template, following instructions about the items that should be included in their request. However, occasionally, due to the lack of a standardized application form the requests have to be returned to the researcher in order to complete any missing information. To solve this problem, a standardized form is being developed. It will contain specific fields to be filled in with the information needed for granting access. This form will be provided via the website, enabling the proposal to be submitted electronically, instead of using a paper form. There is an expectation that this improvement will reduce both the number of returns and the time lag between the data request and approval.

11. Furthermore, a user's guide is being prepared. It will be made available on the IBGE website for download, in order to provide users with knowledge of all the procedures and rules to be followed in applying for access to the Data Laboratory. This initiative aims to ensure transparency to the public and to avoid accusations of secrecy. IBGE considers that it is a good practice to outline that one of the uses of the data from some statistical collections will be to provide researcher access to confidentialised micro-data under controlled conditions for specific purposes.

B. Proposal review

12. The Committee of Assessment of Restricted Data Access is responsible for evaluating the project, based on information provided by the thematic area responsible for the survey micro-data. The Committee authorizes (or not) the access to the data files at Data Laboratory after ensuring that there will be no disclosure risk.

13. The Committee is chaired by the Deputy Director for Surveys and composed by senior staff members from business surveys, methodology and dissemination areas.

14. In order to improve this step of the process, a standardized form to review the proposals is under construction in order to generate standardized assessments among the different thematic areas. Furthermore, an electronic system to manage the whole process is under development. It will replace the current research proposal management which is still done on paper.

15. The use of external data files to match with the requested micro-data files is usually allowed unless a disclosure risk is observed. Currently, this assessment is made by the thematic analyst on a non-structured basis, before the data merging. The idea is to be more rigorous and careful with the data matching in a way that the disclosure risk is decreased.

16. Another initiative to make this matching more restrictive is to require the researchers to obtain a signed formal agreement by the data source managers, for ensuring that the data they provided had been consented for merging.

C. Formal agreements to access

17. Once a project has been approved, formal agreements are established between the researcher and the agency. These agreements outline the conditions of access and the fees for using the Data Laboratory. It is under evaluation whether the researchers should be associated with a university or a research institute in order to access micro-data. In that case the organization would also have to sign the agreement and be responsible for its staff to follow the access rules. Currently, there is no restriction for an independent researcher to gain access to micro-data for scientific use in IBGE.

D. On-site access

18. The databases are installed in the room with special computers for the researchers. The security features of the computers for this use include blocking the network from transferring data and disabling external disk drives and serial parallel ports. The identification code of enterprises has been recoded in the databases from business surveys of IBGE or external sources.

19. After the researchers conclude their work, they save the output in the hard disk of the computer and then prepare a report. A CD-ROM with this information is prepared by IBGE staff in order to be analyzed by the thematic survey area. To improve this procedure, the use of a secure internal network to transfer the data files is under development. This will reduce the disclosure risk in case of loss of the CD-ROM.

20. Another improvement to be implemented is that only those data items required and that are actually needed to conduct the proposed analyses will be included in the analytic data file, on the “need to know” basis. Today, all the variables of the survey are included in the file offered to the researcher.

E. Output checking and release

21. The statistical output must be analyzed before releasing it to the researcher to assure the technical assessment of disclosure risk and confidentiality requirements. The output checking is made by the thematic area responsible for the survey micro-data. Once the output of the project has been approved, i.e., the thematic area assesses that there is no risk of disclosure, another formal agreement is established.

22. This agreement form outlines the conditions of use of the data generated by the special access, where the user has to recognize that data are property of IBGE and has to inform this special access when releasing the results and analysis involving these data.

23. In order to improve the disclosure risk assessment, a technical group composed by statisticians has been conducting in-depth studies on Statistical Disclosure Methods to implement these methods on all the data files and tables produced by IBGE, including the micro-data restricted files used by researchers. Currently, the procedure for anonymizing consists only of removing the identification variables and making cell aggregation or suppression to prevent disclosure.

III. Future work

24. On the “Guidelines and core principles for managing statistical confidentiality and micro-data access” the Principle 4 recommends: “The processes for researcher access to

micro-data as well as the uses and users of micro-data should be transparent, and publicly available.”

25. The IBGE web site is an effective way for providing information on how to grant access to researchers. Currently, the procedures for submitting research proposals are only provided through intranet and the users are informed about these procedures only when they ask for special data.

26. Considering this, IBGE will be improving publicity and transparency about the restricted data access, outlining that data will be provided to researchers under controlled conditions for specific research purposes. Better visibility of these arrangements is necessary to increase public confidence that micro-data will be used properly. The idea is to be completely transparent about the specific uses of micro-data to avoid suspicions of misuse and to ensure that researchers are aware of the consequences for them and for their institution if there are breaches of the agreements.

IV. Concluding observations

27. It is essential to try to meet the needs of the research community while maintaining confidentiality and avoiding disclosure risk. To provide restricted data access for research purposes requires collaboration between all people involved and preparation to deal with a variety of challenging situations. Despite the limitations, IBGE is looking forward to improve the restricted data access procedures and to meet the needs of the research community while protecting confidentiality.

28. In this sense, IBGE is striving to get more information on the experience of other national statistical offices, including the study of Statistical Disclosure Methods. It will adopt the guidelines and core principles and prepare to move from risk avoidance to risk management in providing micro-data access.

V. References

Bianchini, Z.M.; Albieri, S. *Confidentiality and disclosure control: the restricted data access at the Brazilian Institute of Geography and Statistics*. Paper presented at European Conference on Quality in Survey Statistics (Q2006). Cardiff, UK, 2006.

Guidelines and Core Principles for Managing Statistical Confidentiality and Micro-data Access. Principles and guidelines of good practice – interim guidelines. Submitted by the Task Force on Managing Confidentiality and Micro-data Access. Conference of European Statisticians. Fifty-fourth Plenary Session. ECE/CES/2006/6.

Koeller, P.; Vilhena, F.; Zacharias, M.L.B. *Disponibilização de Acesso a Microdados em Institutos Nacionais de Estatística*. Textos para Discussão. IBGE. Rio de Janeiro. 2013. 20p.

Managing Statistical Confidentiality and Micro-data Access – Draft principles and guidelines of good practice. Joint UNECE/Eurostat work session on statistical data confidentiality. Geneva, Switzerland, 9-11 November 2005.