

FOR PARTICIPANTS ONLY

REFERENCE DOCUMENT

DDR/15

15 March 2001

ENGLISH

ORIGINAL: PORTUGUESE

ECLAC

Economic Commission for Latin America and the Caribbean

First meeting of the Statistical Conference of the Americas of the
Economic Commission for Latin America and the Caribbean

Santiago, Chile, 9-11 May 2001

CHALLENGES AND TRENDS IN MODERNIZATION OF NATIONAL STATISTICAL SYSTEMS

Maintaining the confidentiality of statistical information at the Brazilian
Geographical and Statistical Institute

This document was prepared by Zelia Magalhaes Bianchini, Deputy Director for Research and Coordinator for the Confidentiality Group of the Brazilian Geographical and Statistical Institute and has been reproduced without formal editing. The views expressed herein are those of the author and do not necessarily reflect those of the Organization.

01-2-1811.

Introduction

The **Brazilian Geographical and Statistical Institute(IBGE)** has always had and still has the objective of providing relevant and sufficient information to the public. In order to meet this objective, data has to be provided in as much quantity and detail as resources permit. Increasing the quantity and detail in turn brings a greater risk that one or more of the data produced and disseminated may lead to the disclosure of data on identified or identifiable individuals, entities and corporations. In this context, the public right to know and the **rights** of individuals, entities and corporations to the privacy of personal information come into play.

Technological progress offers opportunities to store, process, access and analyze large bodies of data in an efficient manner, and this only increases users' demands for ever more detailed information. On the other hand, it also offers easier access to identifiable and confidential data, thus increasing the risk of individual disclosure. The **legislation** which gives IBGE the mandate to produce statistics establishes explicitly the obligation for IBGE to provide information, but also imposes the counterpart duty of ensuring that the information received will be treated as confidential, will be used exclusively for **statistical purposes** and may not be used for the purposes of certification, nor serve as evidence in administrative, fiscal, or legal proceedings. (see **Law 5534, of 14/11/68** and **Decree Law 161, of 13/02/67**).

It is not an easy task to strike a balance between what is requested and the degree of access that is appropriate.

In view of the resulting need to establish **policies and procedures** to ensure confidentiality of the information collected, produced, disseminated and stored by IBGE, a **Confidentiality Group** was set up in 1999, which aims to focus on recommending courses of action **to reduce the risk of disclosure and to observe compliance with the requirement relating to the privacy of confidential information**, by ensuring anonymity and minimizing the risks of disclosure.

The Confidentiality Group has tried to operate in a manner that is transparent and visible to all, and above all has tried to ensure the responsibility and commitment of all those participating in the process of information production or who may have been authorized to have access to individual data. This action is supported by legislation, the practice of the vast majority of countries, the recommendations of the International Statistical Institute, and IBGE's own experience of more than fifty years of producing high-quality official statistics, and is **essential for IBGE's continued existence as a public institution worthy of public confidence** and capable of offering impartial and high-quality services with integrity.

The integrity of official statistics and public confidence in them are essential for the good governance of a nation. For this public confidence in official statistics to exist, it is essential to ensure the **integrity of the institutions** responsible for their production, and IBGE is thus critically dependent on public confidence to obtain the information it needs to provide the government and the public with the official statistics needed for informed debate and decision-

making.

One of the essential conditions for maintaining the integrity of and public confidence in a statistical institution such as IBGE (and all of its counterparts around the world) is the **strict observance of the confidentiality of the individual or identified information** which it uses to produce statistics. By definition, statistics are constructions based on the collection of individual information, removing its individuality and identity, and constructing summaries of the relevant characteristics of the set of individuals, corporations, entities, products etc.

There are various aspects to consider when dealing with the confidentiality of information, including the legal, political and ethical aspects as well as the technical and operational dimension. It is dangerous to omit any of these aspects, as the omission may lead to decisions with disastrous effects for the credibility of the institution.

The present text offers a description of the main aspects analyzed in relation to the activities carried out by the Confidentiality Group, which includes future challenges and perspectives.¹ It presents the areas of work, details of the implementation (entities and actors which participate in information production) and the current situation in each area (methods, commitments, responsibilities, paths of access to the information, statistical and ethical law). Many activities are mentioned: research, study, raising public awareness, diagnosis, review of legislation, establishing guidelines, documentation, streamlining procedures, orientation, recommendations, supervision, preparation of standards, etc.

Taking those factors into account, and recognizing the importance of analyzing this issue in depth with the aim of formulating standards and procedures designed to ensure the confidentiality of the information gathered, stored, produced and disseminated by IBGE, the following sections will deal with different aspects of the issue.

Section 2 presents the areas of work which have been adopted and presents a flow chart with an **operational model** which identifies the entities and agents involved at each point of the information production process, and what should be observed at each of these points in relation to confidentiality. This flow chart should serve as a guide for preparing recommendations for establishing an information confidentiality policy for IBGE.

Section 3 presents, in a rather summarized form, the main aspects relating to the **implementation of this operational model** through the basic concepts, the methods and techniques for protection of information, the relevant legislation and the procedures for ethical and moral conduct. It also presents a proposal for controlling access to the institutional data bank.

Section 4 indicates the **challenges and perspectives for the future**, and finally, in **section 5**, presents a vast bibliography on managing the confidentiality of information.

¹ See Bianchini *et al.*, 1999.

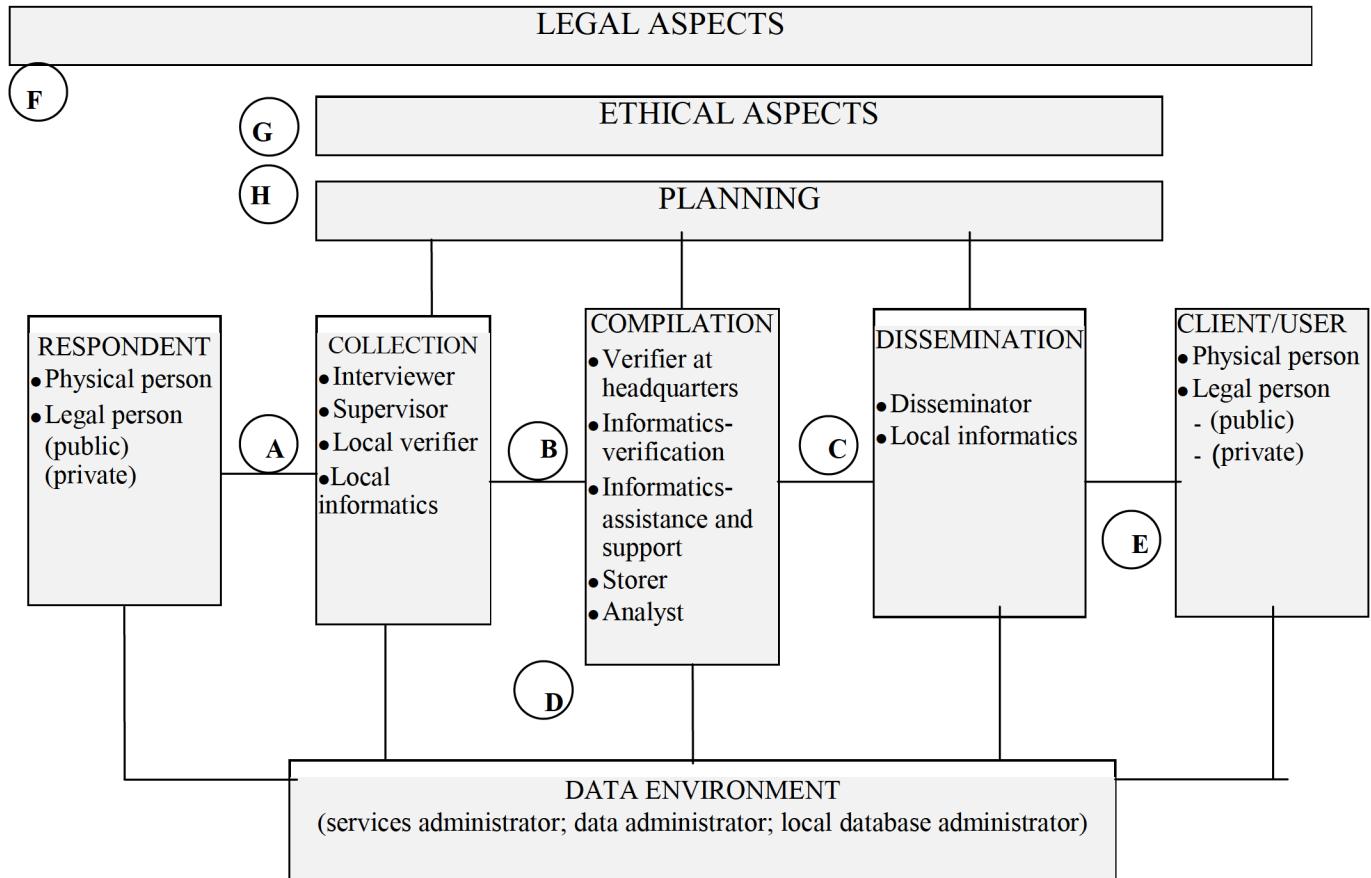
2. Areas of work and operational model

The **main areas of work** identified in the activities carried out are listed below.

- **Bibliographic research on international practices** of the main organizations producing information for dealing with the issue of confidentiality, as a complement to the drafting of recommendations for dealing with this issue in IBGE.
- **Drafting of a national statistical law**, through the revision, updating, extension and **consolidation** of the current legislation, in order to make it democratic in nature and to update the law to suit the current technological age.
- **Conceptual definition and development of appropriate methods and techniques** for minimizing the risks of disclosing confidential **information** in the data prepared for dissemination, whether aggregated or tabulated, microdata, data from property registers, geo-referenced **data**, etc.
- **Definition of the technical and operational aspects** which will be included in the manuals, in order to ensure uniform **treatment** of the confidentiality issue in **the** different areas of IBGE, not only in relation to the media and techniques employed but also the results obtained.
- **Routine granting of access to the store of data** not prepared for dissemination. All access **should** be authorized or denied in accordance with clearly-established criteria, which apply to all officials (whether currently employed or retired) or **persons** connected with external entities with which IBGE has agreements and/or contracts for the provision of services.
- **Definition of the ethical principles** which should inspire the attitude or conduct of officials. These principles should cover the following three aspects:
 1. Compliance with statistical law;
 2. Compliance with existing technical and operational procedures; and
 3. How to deal with cases of negligence.
- **Dissemination of the subject** through seminars to increase awareness, the establishment of an electronic address for comments and suggestions from parties **interested** in contributing to the work of the group, and presentation of the minutes of meetings and the bibliographic references available in digital form on the Intranet.

In order to synthesize and guide the work in these areas, and to facilitate planning of the work, the communication between members and between the latter and the other officials of the institution who participate in the collection, production, storage and dissemination of the information, an **operational model** has been developed, as shown below. For more details, see Anzanello (1999).

OPERATIONAL MODEL AND CONFIDENTIALITY OF INFORMATION



This operational model offers an integrated view of the different aspects of the confidentiality issue (legal, ethical and operational), **in an attempt to identify the relationship between the different entities and agents involved** which act at each point of the information production process and what should be observed at each one of them in relation to confidentiality.

The issue of data storage is related to the various stages of the flow of information, from compilation, through verification and dissemination, until it reaches the user. There are various questions to be answered:

-
- How will the data be stored?
- Who is given access to them?

- What form of access is given?
- How should access and use be controlled?

The issue of security also applies to all stages of the process. There is no point, for example, in having maximum security with regard to the data environment for verification and dissemination, if the same security does not exist at the collection stage.

In the operational model, the text contained in the circles indicate **the points in the information flow at which the issue of confidentiality** should be taken into account. At each observation point, the following are specified:

- The relationship between the entities of the model;
- The agents involved at each point of the information flow;
- What has to be observed (items relating to issues of information confidentiality); and
- Guidelines (for each item considered).

For each of these points we should try to respond to some basic questions relating to the confidential nature of the information:

- (1) Who usually has access to the secret or confidential information and who are their collaborators?
- (2) In these relationships, how can there be a guarantee that there will be no violation of confidentiality?
- (3) Who authorizes people to deal with confidential information which is in printed form? Are there instructions on how to proceed? Is access controlled? Are there sanctions if confidentiality is violated?
- (4) Who authorizes people to deal with confidential processed information that has been processed and stored in digital form? Are there instructions on how to proceed? Is access controlled? Are there sanctions if confidentiality is violated?
- (5) How are the data dealt with when contained in printed form (handling, storage, retrieval, processing, transport, discarding)? Are there instructions on how to proceed?
- (6) How are the data dealt with when they are stored in digital form (handling, storage, copying, retrieval, processing, transmission, discarding)? Are there

instructions on how to proceed?

In addition, the courses of action to be followed in order to resolve individual cases at each of the observation points are defined.

3. Application of the operational model

This section outlines the main aspects considered.

With regard to definitions and methods for preventing disclosure (violation of confidentiality) of data prepared for dissemination, the following text was used as basic reference material: *Statistical Policy Working Paper 22 - Report on Statistical Disclosure Limitation Methodology*, United States, 1994, prepared by the *Subcommittee on Disclosure Limitation Methodology* and coordinated by Nancy J. Kirkendall.

3.1 Basic terms

Disclosure and the various types of disclosure are defined: **disclosure of identity**, disclosure of attributes and deductive disclosure.

Identity disclosure occurs when a third party can identify a respondent on the basis of the data published. In the area of microdata, identification is generally considered as disclosure. For this reason, the methods for limiting disclosure which apply to microdata limit or modify the information which could be used to identify specific respondents or reference units.

Attribute disclosure occurs when confidential information about an individual is disclosed and can be attributed to the latter. The disclosure of attributes presupposes that there has been disclosure of identity. The techniques for limiting disclosure that apply to tables ensure that the respondent's data are published only as part of an aggregate, and with a sufficient number of other respondents to prevent the disclosure of attributes.

Inferential disclosure occurs when the information can be inferred with high reliability on the basis of the statistical properties of the data divulged. For example, the data may show a high correlation between household income and expenditure. If the latter is public information, a third party may use this information to infer expenditure. Such inferences, however, tend to predict aggregate behaviour, rather than individual attributes.

A technique has to be developed, on the basis of classifiers and attributes, to prevent identity disclosure in the various forms of access to data: basic tabulations, special tabulations, microdata from household surveys, assisted access (economic microdata) Internet, etc.

3.2 Methods and techniques

3.2.1 Protection of tabulated data

The selection of a rule to ensure statistical confidentiality for the data presented in tables (tabulated data) depends on whether the data are frequencies or other relevant quantities that should measure something other than a counting of units in a cell of the table.

In the case of **frequency tables**, the main methods for protecting confidentiality are as follows:

- **Sampling and estimation of frequencies in the cells**
- **Special rules** - impose restrictions on the level of detail that can be given in a table, for example, prohibiting tabulations in which a particular cell is equal to the marginal total;
- **The threshold rule** - a cell in the table of frequencies is defined as sensitive if the number of respondents is less than a particular specified number. In order to reduce the risk of disclosure, the tables may be restructured and the categories combined or methods may be used such as cell suppression, random rounding, controlled rounding or confidentiality editing:
 - **Cell suppression** - in a line or column with a suppressed sensitive cell, at least one other cell should be suppressed (complementary suppression);
 - **Random rounding** - instead of using standard rounding conventions, a random decision is made as to whether the value should be rounded up or down;
 - **Controlled rounding** - is a form of random rounding, but with the constraint that the sum of the entries appearing in each line or column have to be equal to the corresponding marginal total;
 - **Confidentiality editing** - a technique for limiting statistical disclosure which can be applied to microdata files before they are used to prepare tables.

The problem of unique disclosure has to be considered in the case of tables which contain magnitude data. The distributions of these values are usually skewed, with a few units having very high values. Disclosure limitation in this case concentrates on making sure that published data cannot be used to estimate accurately the values of the largest units. Protecting the largest values in effect protects all of the values.

For tables of magnitude data, rules have been developed which are called primary suppression rules or linear sensitivity measures, to determine whether a given table cell could reveal individual respondent information. Such a cell is called a sensitive cell and may not be

published. Once the sensitive cells have been identified, there are two options: to restructure the table and collapse cells until no risk cells remain, or to suppress cells.

A number of organizations have found an administrative solution to avoid cell suppression. They obtain written permission to publish a sensitive cell from the respondents that contribute to them. This permission is recognized as a waiver of the promise to protect the confidentiality of the individual information.

The **current practice in IBGE** for dealing with the non-identification of data tabulated for publication usually only applies to economic surveys. The method used is the threshold rule, whereby at least three respondents per cell are required, bearing in mind that:

- only economic surveys referring to manufacturing, trade, services and stocks (in agriculture) adopt this practice of not entering information for cells with less than three respondents, using specific techniques for this purpose;
- traditionally the procedures to prevent identification in this area have been used for the publication of economic census results; the procedure of eliminating information from cells with less than three respondents (substituting an X) has been used; the annual surveys have not used the same procedure, as in view of their non-sample design, they have only published information in more aggregated spatial levels and activity classifications, without however entirely resolving the problem of cells with less than three respondents;
- the threshold method is generally used at present both in the annual industrial survey (PIA) and in the annual trade survey (PAC), requiring at least three respondents per cell and giving preference to the procedure of aggregating levels of classification of activities (or spatial levels) in order to achieve the objective of suppressing less information; there are two ways that these rules can be applied: in industrial surveys the aggregations within the groups of the national economic activities classification (CNAE) follows the criterion of selection of a class/group/division to which the cell with less than three respondents is aggregated as a function of the lower industrial transformation value (VTI). The industrial classification has more entries and thus its subdivisions are more homogeneous; in trade surveys, the aggregation within the CNAE groupings follows the criterion of homogeneity among the activities, as the trade classification has less entries of two and three digits, which makes it more heterogeneous internally;
- in relation to preventing identification in economic surveys, there are two time perspectives to consider, the short-term, with the method already referred to, by carrying out the work manually, and in the medium-term, by using computer programmes to ease the work of preventing identification.

The problem with current procedures for manual processing is the enormous volume of work and time, which can make it impossible to deal extensively and rapidly with the special tabulations in economic surveys. It should be borne in mind that in these surveys are the only form of assistance available to the users, as there are no microdata for these surveys, unlike the case of household surveys.

3.2.2. Protection of published microdata

On the one hand, there is a real demand for microdata with detailed geographical information, and on the other hand, the mandate to ensure the confidentiality of individual information held by IBGE imposes restrictions on data, either by aggregation, suppression or modification, or by controlled or restricted access, or by restriction of the connections. The challenge is to try and offer services which involve microdata without running the risk of one or more of the data produced leading to the disclosure of information on identified or identifiable individuals, entities, and corporations.

The type of data contained in **the microdata file may be classified as identifiers, attributes in the public domain (classifiers) and confidential**, which are differentiated according to the type of unit surveyed: household and/or person, economic unit (corporation, establishment, etc.), product and price.

Some microdata include explicit identifiers, such as name and address. The removal of such identifiers is obviously the first preparatory phase for release of a file where the confidential nature of the individual information has to be protected.

There are also two factors which increase the risks of disclosure. The first is the existence of **high-visibility information**. Some records may contain information on respondents with non-local unique characteristics, such as particular occupations in certain towns (judge, priest, doctor, etc.) or with very high incomes. Secondly, **the microdata file could be superimposed with other files** containing more detailed information.

It is very difficult to prevent disclosure in connection with a microdata file owing to the possibility of **association with external data sources** (Bethlehem, Keller and Pannekoek, 1990). In addition, there are no acceptable levels of disclosure for a microdata file and thus there is no standard which can be applied to guarantee adequate protection.

The methods that reduce the potential for disclosure limit or modify the information. In general, microdata files for public use:

- include only the data from a sample of the population;
- do not include obvious identifiers;
- contain limited geographical detail; and
- contain a limited number of variables.

Additional methods which are used to conceal the more obvious variables include:

- excessive or insufficient coding;
- re-coding into intervals or by rounding;
- addition of or multiplication by random numbers (noise);
- swapping or rank swapping (also referred to as “switching”);
- random selection of records, blanking out selected variables and imputing them again;
- aggregation of small groups of respondents, substituting the average value for the group for the value provided by each individual (also referred to as “blurring”).

IBGE's current practice for disseminating microdata from household surveys is to publish files for public use without identifiers, with little geographical detail, but without any processing to conceal the most obvious variables.

In economic surveys, the main international statistical organizations do not release microdata files. The risks of disclosure in microdata from economic surveys are much higher than for microdata from household surveys, as there is a greater number of variables for identification.

Economic information is usually in the form of magnitude data. These data are usually skewed and allow the respondents to be easily identified from other information available to the public, and may also contain strategic information. For these reasons, economic microdata are not disseminated for public use.

In the face of the high demand for access to the microdata from the agricultural census, IBGE is dealing with requests from various organizations on a case-by-case basis, authorizing on-site access. Access to the microdata is granted in the department of agriculture on the following conditions: presentation of a project, signing an agreement for each specific case which establishes the conditions of use and the responsibility assumed, and explicitly guarantees confidentiality. The main points covered in the agreement for access to such microdata are as follows:

- identification of the project in which they are to be used;
- the timeframe for completion of the project;
- an undertaking to use the microdata only on the premises of IBGE;
- observance of confidentiality in accordance with the relevant legislation;
- an undertaking that the individual information will not be transmitted, sold, or transferred;
- an undertaking to present the results that are to be published;
- if the final product is presented in tabulated form, the procedures to be adopted prior to publication in order to prevent identification.

As a result of this effort to try and meet the new needs of users in the best possible form, while respecting the confidentiality mandate, some guidelines are being developed. One of them

refers to the **creation of facilities and the development of infrastructure and monitoring and follow-up mechanisms**, with an adequate level of security, in order to provide access to the microdata that is identified or easily susceptible to identification. In this regard, Zayatz, Masell and Skel (1999) refer to the creation of special environments, on the part of the Census Bureau of the United States, so that researchers can have access to non-published data.

3.2.3. Protection of microdata with identifiers (property registers)

Confidentiality in relation to property registers is a rather broad issue, owing to the variety of registers available: business register, of agricultural establishments, of homes (referred to in the compilation sheets from the demographic censuses), of products and prices. In the case of registers, the problem of disclosure is one of identification, unlike microdata, when the attributes are associated.

Business register

Firstly, a specific example of an economic register will be considered, namely the **business register - CEMPRE**. The Division of Registers and Classification of the Department of Surveys is the entity responsible for preparing and maintaining CEMPRE, constantly updating it with the results from the surveys and the information from administrative registers produced externally.

In the case of the **business register**, the issue of confidentiality is currently dealt with as follows:

- Confidentiality concerns arise when information from the register is being supplied together with identifying names, addresses, and characteristics of the corporations.
- IBGE supplies the register information selectively, only providing the name, address, activity and age groups of employees to users who can justify their use. It is not provided for commercial purposes, such as a postal catalogue.
- There is a consensus that registers with characteristics of corporations (for example: employees and income) may be provided to regional statistical organizations (OREs) for the purpose of selecting samples, subject to a prior agreement to maintain confidentiality.
- The register does not have a commercial purpose, and the information is not statistical, as there is no aggregation of data.
- The main function of the register is to serve as a support tool for planning economic surveys. The product of IBGE is in the results of its surveys. Another function of the register is to be used as a statistical coordination tool. It is not a question of offering the register to public or private corporations, but of limiting its use to the design of surveys.

- There are other aspects relating to the quality of the information on the register, which could bring into question the credibility of the statistical organization. For example, Statistics Canada does not part with information from the registry, bearing in mind the quality of this information.
- The statistical organizations of other countries, such as Australia, which previously did provide individual data from the register, now announce on their Web pages that they do not provide individual data from the register, but only tabulations. On the other hand, the National Institute of Statistics and Economic Studies (INSEE) of France and the National Institute of Statistics (ISTAT) of Italy do not have restrictions on providing access to registers.

Register of households, agricultural establishments and segments of public areas

Unlike CEMPRE, the **registers of households and agricultural establishments** are not yet structured in a digital format. The main source of information for these registers are the forms that are filled out for the demographic census² and the agricultural census respectively, which are inserted in the census booklets. Also, a listing operation is carried out for each household survey, whereby a register is prepared for all sectors selected in each sample, in order to select the households.

IBGE responds to requests for copies of the forms or the listing by providing the address and removing the name and all other attributes.

A **register of segments of public areas** is being drawn up by sector, using information from the forms for the 1996 population count: the names of the public areas contained in the sector, the number of units visited and of individual households. This register was extremely useful for revising the geographical basis for the 2000 census. This register will be updated with the compilation sheets from the 2000 census.

Register of products and prices

Traditionally, IBGE only provides **average prices** for generic products, without identifying the brand. Despite the strong demand for providing the average prices of specific products according to brand name, IBGE does not provide this information, in order to ensure confidentiality, thus protecting the respondent and preserving the reputation of the institution.

IBGE's difficulty in meeting these requests is related to the definition of a statistic and how information compiled by IBGE is related to that process, as it is not a question of anonymous averages but of averages referring to products and brand names that are clearly specified and thus identified. This association between the data and the identity is what causes the problem, and is the reason that IBGE refuses requests of this nature.

² It is planned to transcribe the 2000 demographic census information into digital format.

3.2.4. Protection of geo-referenced data

The group examined the issue of statistical disclosure in relation to **small areas**, based on the articles of Rees and Duke-Williams (1997 and 1998), prepared as part of the project for the control of statistical disclosure, with resources from the European Union's ESPRIT programme.

In order to prevent disclosure, one of the protection measures applied is the definition of a minimum number of persons or households for generating the statistics in different areas. If the statistics are generated for two geographical bases, then operations of the system of geo-referenced information (SIG) can be used to subtract one set of tables from the other in order to obtain statistics for areas that are below the confidentiality limit. The danger of differentiation occurs when it is possible to nest small areas from one geographical base into areas from another geographical base.

These experiences have suggested that the practical risks are much less than those suggested in theory and that it is safe to publish statistics for secondary geographical bases if they are large enough in relation to the primary areas.

The recommendations of Rees and Duke-Williams (1997) for preventing the problem of disclosure are as follows:

- Publish statistics from small areas for the ten-year censuses. There is a high demand for this and it is a more cost-effective strategy than offering tables on request.
- Choose a strategy for the production of safe tables. use the minimum levels of the confidentiality threshold for the tables.
- Adapt the primary geographical base selected for the most general possible use.
- The risk of publishing in alternative geographical bases whose zones are greater than the primary census units is negligible.
- It is in any case not safe to produce a secondary set of statistics on small areas from another very small geographical base which is relatively close to the primary geographical base selected. This is true for the publications of current, past or future census statistics for small areas.
- When time series are being constructed from statistics from small areas, it is more useful to extend the current geographical base back into the past than to try to preserve the past geographical basis. A survey which considers the future of the tables for consultation makes it possible to construct such time series.

IBGE, when it has concluded the system of aggregation of sectors, for forming weighting areas from the 2000 census sample, will have to consider the possibility of making it compatible

(by aggregation) with the postal base which is currently being revised by the postal service. The aggregation of sectors reduces the confidentiality problem but does not eliminate it.

3.3 Ethical and moral conduct and procedures

A **manual on ethical and moral conduct** is being developed, and is already available in a preliminary version, intended for IBGE employees, and based on various legal tests aimed at civil servants in general, and formal texts relating to the generation of official statistical and geo-scientific information, including the *Principios fundamentales de las estadísticas oficiales* (28th session of the United Nations Statistical Commission, April 1994).

Whereas the technical and operational guidelines are objective in nature, the precepts of ethical and moral conduct are subjective; the guidelines will thus be more effective than the precepts, although it is still crucially important that the latter be clearly stated.

It should be noted that **ethics** refers to the matter of applying philosophical principles to everyday life, of reflecting on the reasons for desiring justice and harmony, and the means for achieving them, but above all it is concerned with a theory of values; and **morals** refers to the specific customs, values and standards of behaviour of a particular society, and refers to the construction of a set of prescriptions designed to ensure a just and harmonious community life. That is, if **ethics** on the one hand considers human behaviour from a valuative and normative point of view, treating it in a generic and abstract manner, **morals** on the other hand establishes specific principles which have to be followed in order to have a community life.

In general, in the information generating process there are standards of conduct to be observed at the stages of planning, design, processing, analysis and interpretation, data bank maintenance, and dissemination, as well as other aspects.

Taking into account the above, it is important to be clear that a **manual of ethical and moral conduct** acts in three complementary ways, namely: firstly, by requiring and offering guidance for rigorous and unconditional compliance with the statistical law; second, by requiring and offering guidance for strict and rigorous compliance with the technical and operational guidelines; third, by indicating how to deal with cases of negligence and omission in the former areas. By focusing on these three aspects, the manual concentrates on action which is persuasive and educational (continuing education) in nature, while also stating the penalties to be applied in cases of contempt.

Also, it is appropriate to guide the institutional structure of the whole process by establishing principles for the creation of an **ethics committee**, as a guarantee of the democratic good faith of the disciplinary processes.

As for the **technical and operational manual**, aspects related to the use of certain computer resources were analyzed, mainly those relating to the use of electronic mail and the Internet. In many cases, it is difficult to separate the operational aspect from the legal and moral

aspects. The two topics dealt with (electronic mail and the Internet) are very clear examples of the difficulty mentioned to the extent that objective rules on how to use these resources with care are mixed with the legal issues (limitations on the monitoring of mail messages and attached files) and the moral issues (right to privacy and monitoring access to data) which arise.

3.4. Access to the institutional data bank

This section presents a proposal that has been analyzed by the Confidentiality Group on access to the institutional data bank in relation to the **data environment** of the operating model (text D).

Persons, whether they are officials of IBGE or of entities external to IBGE, may have access to the institutional store of microdata of IBGE through an authorization which is granted by those responsible for the different units of IBGE, on the basis of rules defined by the institution.

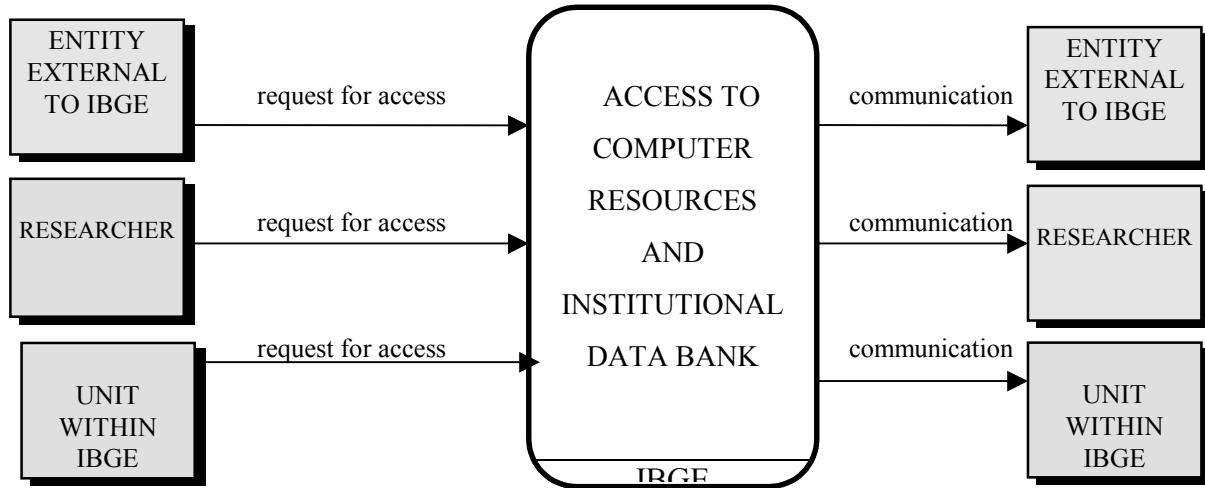
In order to have access to the store of data which have not been prepared for dissemination, prior access is required to IBGE computer resources. For this purpose, whoever authorizes access to the store also authorizes access to the computer resources that are needed to facilitate access to the data store.

When such authorization has been given, the granting or restriction of access to resources and data is arranged by the IBGE Department of Informatics.

Two diagrams are presented below which help to visualise the main entities and flows of information in the process of authorization, registration and control of access in two different levels of detail.

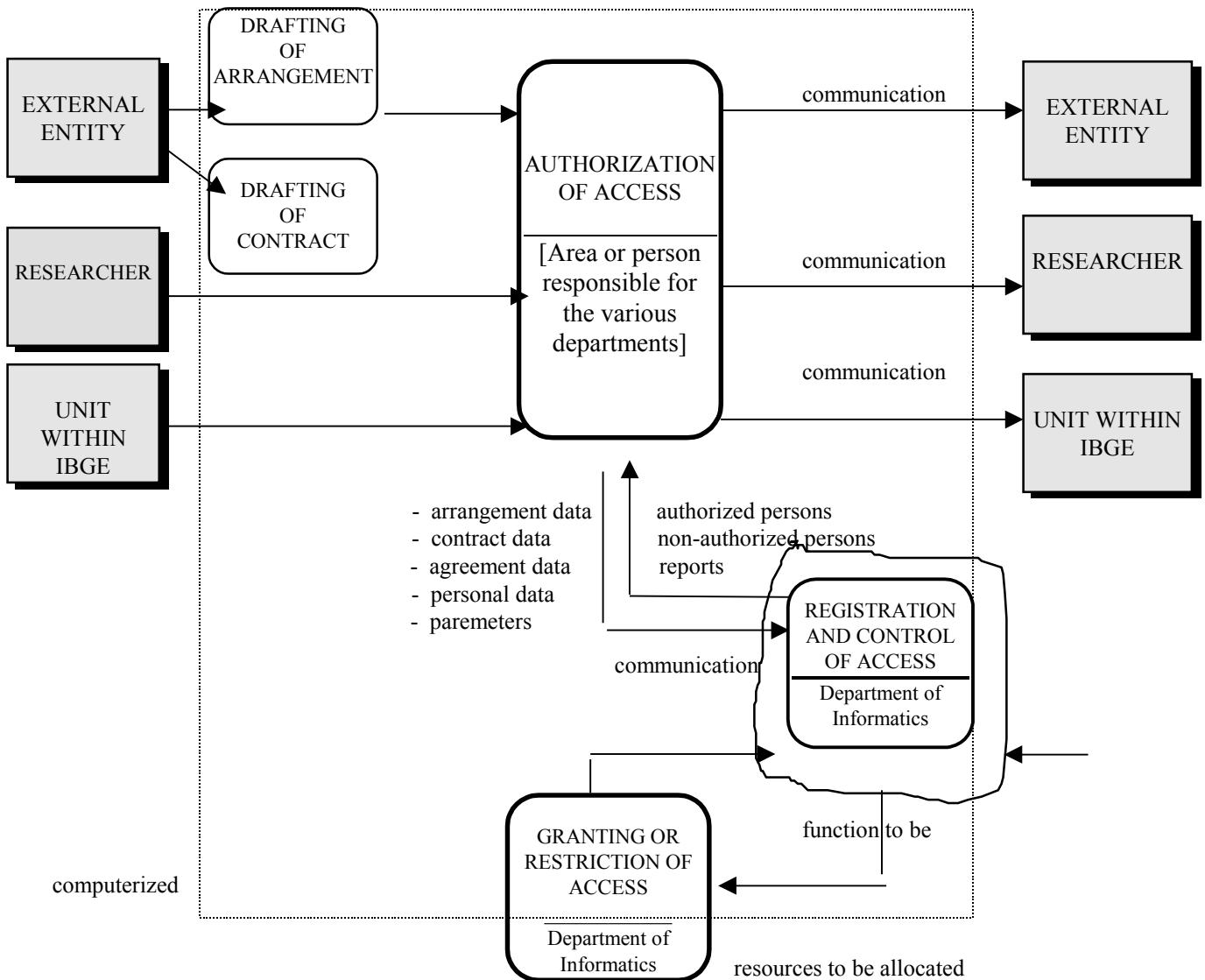
Bianchini and others (1999) contains a detailed description of the proposal for the conditions for authorizing or denying access to the data, which apply to all officials, currently employed or retired, persons related to external entities, physical persons, whether or not related to legal persons, by arrangement or agreement. They also establish the competencies relating to access to computer resources and data from the institutional store, with specification of the data to be included or removed from the register.

PROCESS OF AUTHORIZATION, REGISTRATION AND CONTROL OF ACCESS CONTEXTUAL DIAGRAM



- authorized persons
 - non-authorized persons
 - monitoring reports
 - arrangement data
 - contract data
 - agreement data
 - personal data
 - parameters

**ACCESS TO IBGE COMPUTER RESOURCES AND MICRODATA STORE PROCESS OF AUTHORIZATIION, REGISTRATION AND CONTROL OF ACCESS
DIAGRAM LEVEL 0 (ZERO)**



4. Future challenges and perspectives

In view of the above, it is clear that there has been a lot of progress and new understanding in the area of confidentiality of information, but that there is much more to do, in view of the complexity of the subject. The main tasks include the following:

- The drafting of a **statistical law** which will support the competent authorities, bringing together and updating current legislation, taking into account the current environment of democracy and technology.
- The development of a **technical and operational manual** which contains the rules and procedures to be followed by the various actors at each point of the operating model.
- Completion of the **manual of ethical and moral conduct** which contains the principles to be applied by IBGE officials in their capacity as actors at the various points of the operating model.

In developing these drafts, efforts are made to ensure decentralized implementation by assigning resources to projects for the following activities:

- Research, absorption, verification, evaluation and dissemination of **automated protection methods and systems**. This project should also include the **creation of courses for the annual training programme** on confidentiality and its effects on survey methods and processes.
- Establishing procedures which ensure complete **security of access** to the institutional data bank.
- **Creation of adequate facilities for special access** to the data bank for authorized persons.
- Acquisition of the equipment needed for **appropriate destruction of material** containing individual information.

There are three observations to make with regard to ensuring that confidentiality is guaranteed to the respondents, which is an essential issue for the present and future of statistical organizations, so that more work can be carried out with greater efficiency over the next few years:

- The measures adopted or to be adopted require a significant amount of human resources, but not only this, as the **costs** of this set of activities have to be taken seriously.
- The present and future actions imply changes in the institutional **culture**, which requires action at all the functional levels, without exception, involved in the various stages of generation of the information;

- The **planning of each survey** should include the procedures for information protection, from compilation to dissemination. Commitment is needed from all the relevant departments and areas of the organization, with the help of training, and with a detailed description of the current situation in each of the surveys.

In view of the above, it should be understood that the results, however urgent and essential they may be, will only be felt in the medium and long-term, in view of the time and special delays that unfortunately exist in the institution.

Finally, in view of the fact that the issue requires constant attention, the Committee on Confidentiality was recently established to serve as a place of reflection for issues relating to confidentiality, with the following functions:

- (a) **to propose solutions** for questions relating to confidentiality of information of a statistical and geo-scientific nature, personal or identified, compiled, produced, stored and disseminated by IBGE;
- (b) **to consider** requests for access to confidential information, giving its opinion on the authorization and conditions of use; and
- (c) **to evaluate** systematically compliance with the promise to ensure that the confidential nature of the information is respected.

Bibliography

- Anzanello, E. (1999), *Modelo de operação para pesquisas estatísticas e sigilo de informações*, Rio de Janeiro, Brazilian Geographical and Statistical Institute (IBGE).
- Bethlehem, J.G., W.J. Kellere and J. Pannekoek (1990), “Disclosure control of microdata”, *Journal of the American Statistical Association*, No. 85.
- Bianchini, Z. M. and others (1999), *Grupo de Sigilo: atividades realizadas, desafios e perspectivas para o futuro*, Rio de Janeiro, Brazilian Geographical and Statistical Institute (IBGE).
- Bianchini, Z.M. (1995), *Considerações sobre o sigilo das informações. Mercosur: sinopsis estadística*, vol. 2, Buenos Aires.
- Burshtein, G. and G. Lang (1999), “Administration and policy of statistical data confidentiality in Israel”, *Joint ECE/Eurostat Work Session on Statistical Data Confidentiality*, Working Paper, No. 22, Athens.
- Carvalho, J. (1999), *Sigilo: Legislação básica*, Rio de Janeiro, Brazilian Geographical and Statistical Institute (IBGE)/ General Counsel.
- Cecil, J.S. (1993), “Confidentiality legislation and the United States Federal Statistical System”, *Journal of Official Statistics*, vol. 9, No. 2.
- Cook, L. (1998), *Protocols for Official Statistics*, New Zealand.
- Cox, L.H. (1994), “Protecting confidentiality in establishment surveys”, *Survey Methods for Businesses, Farms and Institutions*, Brenda Cox (ed.), New York, John Wiley and Sons.
- Defays, D. and M.N. Anwar, (1998), “Masking microdata using micro-aggregation”, *Journal of Official Statistics*, vol. 14, No. 4.
- Duncan, G.T. (2000), “Optimal disclosure limitation strategy in statistical databases: deterring tracker attacks through additive noise”, *Journal of the American Statistical Association*, vol. 95, No. 451.
- Elliot, M. and A. Dale (1999), “Scenarios of attack: the data intruder’s perspective on statistical disclosure risk”, *Netherlands Official Statistics*, vol. 14, special issue.
- European Union (1999), *El secreto estadístico en las legislaciones nacionales de los Estados miembros de la Unión Europea*, Madrid.
- Fienberg, S.E. and L.C.R.J. Willenborg (1998), “Introduction to the special issue: disclosure limitation methods for protecting the confidentiality of statistical data”, *Journal of Official Statistics*, vol. 14, No. 4.
- Fienberg, S.E. and U.E. Makov (1998), “Confidentiality, uniqueness, and disclosure limitation for categorical data”, *Journal of Official Statistics*, vol. 14, No. 4.
- Fuller, W.A. (1993), “Masking procedures for disclosure limitation”, *Journal of Official Statistics*, vol. 9, No. 2.
- Gouweleeuw, J.M. and others (1998), “Post randomisation for statistical disclosure control: theory and implementation”, *Journal of Official Statistics*, vol. 14, No. 4.
- Greenberg, B. (1990), “Disclosure avoidance research at the Bureau of the Census”, *Proceedings of Sixth Annual Research Conference*, Washington, D.C., Bureau of the Census.
- GSS (United Kingdom Government Statistical Service) (1995), *Official Statistics: Code of Practice*, London.

- Hurkens, C.A.J. and S.R. Tiourine (1998), "Models and methods for the microdata protection problem", *Journal of Official Statistics*, vol. 14, No. 4.
- IBGE (Brazilian Geographical and Statistical Institute) (1996), *Requisitos técnicos para uma lei sobre a produção e o uso de informações estatísticas*, Rio de Janeiro.
- INDEC (Instituto Nacional de Estadística y Censos) (1988), *Manual sobre secreto estadístico*, vol. 2, Buenos Aires.
- Jabine, T.B. (1993), "Statistical disclosure limitation practices of United States Statistical Agencies", *Journal of Official Statistics*, vol. 9, No. 2.
- Keller, W.J. and J.G. Bethlehem (1992), "Disclosure protection of microdata: problems and solutions", *Statistica Netherlandica*, No. 46.
- Keller-McNulty, S. and E.A. Unger (1998), "Database system prototipe for remote access to information based on confidential data", *Journal of Official Statistics*, vol. 14, No. 4.
- Kirkendall, N.J. and others (1994), "Report on Statistical Limitation Methodology (Subcommittee on Disclosure Limitation Methodology)", Statistical Policy Working Paper, No. 22, Subcommittee on Disclosure Limitation Methodology.
- Kooiman, P., J. Nobele and L. Willlenborg (1999), "Statistical data protection at Statistics Netherlands", *Netherlands Official Statistics*, vol. 14, special issue.
- Lang, G. (1999), "Statistical confidentiality and the French committee of statistical confidentiality concerning enterprises", *Joint ECE/Eurostat Work Session on Statistical Data Confidentiality*, Working Paper, No. 20, Athens.
- Marsh, C., A. Dale and C. J. Skinner (1994), "Safe data versus safe settings: access to microdata from the British Census", *International Statistical Review*, vol. 62, No. 1.
- McLenaghan, J.B. (1999), "The UN's fundamental principles of official statistics issues of covarage and application", *Working Session On Best Practices in Statistics*, Singapore.
- Michael, J.A. and others (1978), "Report on Statistical Disclosure-Avoidance Techniques (Subcommittee on Disclosure-Avoidance Techniques Federal Committee on Statistical Methodology)", Statistical Policy Working Paper, No. 2, Subcommittee on Disclosure Limitation Methodology.
- Nordholt, E.S. (1999), "Statistical disclosure control of Statistical Netherlands employmenent and earnings data", *Netherlands Official Statistics*, vol. 14, special issue.
- Openshaw, S., O. Duke-Williams and P. Rees (1997), "Measuring Confidentiality in Census Data", Working Paper, No. 97/8.
- Pannekoek, J. and T Waal (1998), "Synthetic and combined estimators in statistical disclosure control", *Journal of Official Statistics*, vol. 14, No. 4.
- Rees, P. and O. Duke-Williams (1998), *Data Privacy Considerations: The Differencing Problem in Statistical Disclosure*, London.
- _____, (1997), *Experiments with and Recommendations for the Creation and Release of Small Statistics from National Censuses*, London, University of Leeds.
- Robertson, D.A. (1993), "Cell suppression at Statistics Canada", *Proceedings of the Bureau of the 1993 Census Annual Research Conference*, Washington, D.C., Bureau of the Census.
- Samuels, S.M. (1998), "A Bayesian, species-sampling-inspired approach to the uniques problem in microdata disclosure risk assessment", *Journal of Official Statistics*, vol. 14, No. 4.
- Senra, N.C. (1999a), *Garantía de Sigilo, a deontología do estatístico: sua conduta ético-moral*, Rio de Janeiro, Brazilian Geographical and Statistical Institute (IBGE).

- _____ (1999b), *Reflexões sobre a questão do sigilo*, Rio de Janeiro, Brazilian Geographical and Statistical Institute (IBGE).
- Silva, P.L.N. (1999), *Sigilo Estatístico e disseminação de informações*, Rio de Janeiro, Departament of Methodology, Brazilian Geographical and Statistical Institute (IBGE).
- _____ (1988), *O sigilo das informações estatísticas: idéias para reflexão*, Rio de Janeiro, Brazilian Geographical and Statistical Institute (IBGE).
- Skinner, C.J. and D.J. Holmes (1998), "Estimating the re-identification risk for record in microdata", *Journal of Official Statistics*, vol. 14, No. 4.
- Skinner, C.J. and others (1990), "Disclosure avoidance for census microdata in Great Britain", *Proceedings of the Sixth Annual Research Conference*, Washington, D.C., Bureau of the Census.
- Spieker, F. (1999), "Access to confidential data in an integrated statistical system", *Joint ECE/Eurostat Work Session on Statistical Data Confidentiality*, Working Paper, No. 18, Athens.
- Thorogood, D. (1999a), "Statistical confidentiality and the European level", *Joint ECE/Eurostat Work Session on Statistical Data Confidentiality*, Working Paper, No. 16, Athens.
- _____ (1999b), "Protecting the confidentiality of Eurostat statistical outputs", *Netherlands Official Statistics*, vol. 14, special issue.
- United Nations Statistical Commission (1994), *Fundamental Principles of Official Statistics*, New York.
- Vasconcellos, M.T.L. (1996), *A disseminação dos dados e a proteção do informante: idéias básicas para a definição de uma política institucional*, Rio de Janeiro, Departament of Methodology, Brazilian Geographical and Statistical Institute (IBGE).
- Waal, T. and L. Willenborg (1999a), "Information loss through global recoding and local suppression", *Netherlands Official Statistics*, vol. 14, special issue.
- _____ (1999b), "Exact disclosure in a super-table", *Netherlands Official Statistics*, vol. 14, special issue.
- _____ (1998), "Optimal local suppression in microdata", *Journal of Official Statistics*, vol. 14, No. 4.
- Willenborg, L.C.R.J. (1993), "Discussion: statistical disclosure limitation", *Journal of Official Statistics*, vol. 9, No. 2.
- Willenborg, L.C.R.J. and T. Waal (1996), "Statistical disclosure control in practice", *Lecture Notes in Statistics*, No. 111, New York, Springer-Verlag.
- Willenborg, L.C.R.J., R.J. Mokken and J. Pannekoek (1990), "Microdata and disclosure risks", *Proceedings of the Sixth Annual Research Conference*, Washington, D.C., Bureau of the Census.
- Zaslavsky, A.M. and N.J. Horton (1998), "Balancing disclosure risk against the loss of nonpublication", *Journal of Official Statistics*, vol. 14, No. 4.
- Zayatz, L., P. Massell and P. Steel (1999), "Disclosure limitation practices and research at the U.S. Census Bureau", *Netherlands Official Statistics*, vol. 14, special issue.