Fundação Instituto Brasileiro de Geografia e Estatística - IBGE
Diretoria de Pesquisas - DPE
Divisão de Metodologia - DME

O Tratamento da Questão do Sigilo das Informações

Zélia Magalhães Bianchini

Rio de Janeiro
Janeiro / 1994

APRESENTAÇÃO

Este texto foi preparado com o objetivo de subsidiar o debate interno na Diretoria de Pesquisas, para as atividades preparatórias do I Encontro dos Órgãos Estaduais Produtores de Informações Estatísticas e Cartográficas integrantes do Sistema Estatístico Nacional - SEN, a ser promovido pelo IBGE na segunda quinzena de abril de 1994.

O objetivo desse Encontro é avaliar a atual organização do SEN e estabelecer as bases para um efetivo processo de integração e articulação técnica entre seus membros.

O texto enfoca o tema relacionado com a questão do **sigilo das informações** e baseia-se principalmente nas leituras do texto publicado na Série Textos para Discussão nº 4 de autoria de Pedro Luis do Nascimento Silva e de alguns artigos publicados no volume especial da *Journal of Official Statistics - JOS*, em 1993, que enfoca o tema *Confidentiality and Data Access*.

1. INTRODUÇÃO

A legislação brasileira protege o informante não permitindo que os dados fornecidos sejam usados para qualquer fim tributário ou jurídico. Além disso, os órgãos de estatística são, em geral, obrigados por lei a respeitar a privacidade dos informantes e coletam os dados sob a promessa de confidência. Dessa forma, cabe a esses órgãos a responsabilidade de definir estratégias para a liberação dos dados, de forma que não viole a promessa de não identificação dos dados.

Além do aspecto legal, não há dúvidas de que violar essa promessa pode destruir a credibilidade da instituição para a coleta dos dados e gerar diversos inconvenientes ao informante, como por exemplo municiar sua concorrência de dados estratégicos.

Uma importante área da pesquisa estatística é desenvolver métodos que permitam divulgar o máximo possível dos dados, respeitando o aspecto legal para proteger a privacidade individual prometida aos informantes. A meta de um órgão de estatística é maximizar o nível de informação a ser dada ao público, sujeita ao requerimento de que o risco de revelação seja aceitavelmente baixo.

Um diagnóstico do tratamento dado à questão do sigilo das informações pelo IBGE foi dado em Silva(1988). O capítulo 2 apenas endossa as questões já levantadas e não discutidas ou estudadas no referido artigo.

O capítulo 3 apresenta algumas considerações sobre o tratamento da questão do sigilo por órgãos de estatística internacionais, cujo enfoque principal é mencionar a vasta bibliografia sobre o assunto.

O capítulo 4 contém as considerações finais, onde é reiterada a necessidade de criação de um grupo de trabalho para estudar, disseminar e propor normas e procedimentos para o tratamento da questão do sigilo das informações no IBGE. Por último, no capítulo 5, são listadas as referências utilizadas.

2. O TRATAMENTO DA QUESTÃO DO SIGILO PELO IBGE

A respeito do sigilo das informações já não é de hoje que foi proposta uma discussão para estudar, aqui no IBGE, a questão em profundidade. Em abril de 1988 foi publicado na Série Textos para Discussão um artigo escrito por Pedro Luis do Nascimento Silva, chefe do então Núcleo de Metodologia, intitulado **O Sigilo das Informações Estatísticas - Idéias para Reflexão** (veja Silva 1988).

O referido trabalho foi elaborado com a finalidade de subsidiar o debate interno na Diretoria de Pesquisas, durante as atividades preparatórias do Seminário de Disseminação de Informações, realizado em dezembro de 1987 e da III CONFEST - Conferência Nacional de Estatística, realizada em novembro de 1989.

Silva(1988) aborda diversos aspectos do problema: legal, ético, político e técnico-operacional. Apresenta conceitos e alternativas para a pesquisa, conclusões e recomendações. As principais conclusões merecem ser reproduzidas por serem tão pertinentes e estarem redigidas com tanta propriedade e clareza:

"a) o tratamento da questão do sigilo é precário e heterogêneo nas diversas áreas do IBGE, não só com respeito aos meios e técnicas empregados como também com relação aos resultados que se obtêm; basicamente compreende a repetição de procedimentos tradicionais, que ignoram os avanços tecnológicos, quer na área de metodologia, quer no campo da informática, e que sequer se acham suficientemente documentados, normatizados ou mesmo incorporados à rotina das pesquisas; algumas vezes, são do conhecimento de uns poucos funcionários que têm a seu encargo o trabalho de aplicá-los;

b) o IBGE está bastante atrasado na apreensão e utilização rotineira de métodos e técnicas apropriados para minimizar os riscos de revelação de informações confidenciais, mas este atraso pode ser rapidamente vencido com a alocação dos recursos necessários para o desenvolvimento de um projeto de pesquisa sobre o tema;

c) falta uma conscientização maior dos quadros de gerência da organização para o problema, que se espera desapareça a partir do debate em torno desse documento;

d) há diversos aspectos do problema cuja abordagem terá que passar, necessariamente, por decisões políticas, legais e éticas que devem ser orientadás tecnicamente, mas tomadas no âmbito da direção superior; um exemplo é a liberação de dados para outros órgãos membros do Sistema Estatístico Nacional;

e) não é possível ignorar nem o lado dos órgãos produtores nem o lado dos órgãos disseminadores das informações ao abordar a questão do sigilo - o que se tem é o dilema de dar a mais ampla divulgação aos dados produzidos, visando maximizar a sua utilidade, mantendo em sigilo as informações confidenciais dos indivíduos, sem o que o futuro da produção de informações estará certamente comprometido."

Em função das conclusões mencionadas foi fortemente recomendado, em Silva(1988), a criação de um grupo de trabalho encarregado de propor normas e procedimentos destinados à manutenção do sigilo das informações coletadas, produzidas, disseminadas e armazenadas pelo IBGE.

Apesar das duras conclusões e das recomendações, até onde se sabe pouco ou quase nada foi feito, desde a divulgação do referido TEXTO PARA DISCUSSÃO, em busca da formulação de uma política de disseminação de informações.

Temos novamente uma ocasião propícia para que o assunto seja tratado e aprofundado. Para que o IBGE cumpra o seu papel de coordenador do SEN é urgente a realização do trabalho conjunto de técnicos da área de produção, disseminação e demais áreas relacionadas com a questão, com o objetivo de fornecer elementos que assegurem a normatização dos procedimentos a adotar com vistas ao tratamento do sigilo das informações.

3. O TRATAMENTO DO SIGILO POR ÓRGÃOS DE ESTATÍSTICA INTERNACIONAIS

Nos últimos anos vem ocorrendo uma crise mundial de credibilidade das informações investigadas, como conseqüência do aumento das taxas de não-resposta e do aumento da dificuldade em obter o acesso às informações.

Para estudar os problemas que as instituições oficiais de estatística vinham enfrentando, o *U.S.* Committee on National Statistics, do *U.S. National Academy of Sciences - National Research Council reuniu-se* com a *Social Science Research Council* para conveniar em 1989 o *Panel on Confidentiality and Data Access.* O suporte para esse painel de estudos foi dado por muitos órgãos federais de estatística, tais como: *National Science Fundation, the Bureau of the Census, the Bureau of the Labor Statistics, the Internal Revenue Service Statistics of Income Division, the National Institute on Aging, the National Center for Education Statistics.*

O objetivo do estudo era obter recomendações que viessem a auxiliar os órgãos federais de estatística, tendo sido tratados três aspectos de interesse: proteção dos dados através de procedimentos que assegurassem a privacidade e credibilidade, aumento da confiança e facilidade para disseminação.

Em março de 1991, foi promovida pelo painel de estudos a *Conference on Disclosure Limitation Approaches and Data Accesss.* Seu objetivo foi examinar a situação americana, do ponto de vista da ciência estatística e computacional bem como o aspecto legal, do da ciência social. Os artigos apresentados na conferência constituem a maior parte dos artigos publicados no volume 9 nº 2 de 1993 da *Journal of Official Statistics - JOS,* que trata de um volume especial sobre *Confidentiality and Data Access.*

Em anexo, consta uma cópia do sumário dos 11 artigos e discussões publicados na referida JOS, bem como cópia do seus resumos *(abstract).*

Dentre os 11 artigos destaca-se o de Jabine(1993), que se refere a uma descrição da política, prática e procedimentos usados nos órgãos federais de estatística dos Estados Unidos com relação à questão do sigilo das informações. O artigo apresenta um capítulo que sintetiza o tratamento dado por cada órgão, apresentando modelos de alguns documentos oficiais. À guisa de ilustração, são apresentadas, a seguir, algumas das normas adotadas para o tratamento da questão do sigilo das informações.

- Economic *Research Service (ERS)* e *National Agricultural Statistics Service (NASS)* do *Department of Agriculture*

Para pesquisas amostrais as estimativas para serem publicadas devem obedecer aos seguintes critérios: devem ser disponíveis no mínimo 3 observações por célula; o dado sem expansão de algum informante deve representar menos que 60% do total que está sendo publicado, exceto quando tiver a permissão por escrito do informante.

- *Bureau of the Census*

Pioneiro na divulgação de micro-dados, o *Bureau* desenvolveu procedimentos específicos para assegurar o sigilo das informações para essa modalidade de disseminação dos dados. Em 1981 foi criado o *Microdata Review Board*. Um dos critérios adotados por esse comitê foi o de que as áreas geográficas com menos de 100.000 habitantes não poderiam ser identificadas nos arquivos contendo micro-dados. Esse corte foi adotado a partir de 1981, sendo que anteriormente o limite era de 250.000 habitantes.

- *Bureau of Labor Statistics (BLS)*

As técnicas usadas para limitação da revelação estatística no BLS variam por programa. As mais comumente usadas são as a que se referem à medida de concentração da informação, variando de 30% a 80% dependendo da pesquisa, e as que requerem um número mínimo de informantes ou observações por célula.

Para divulgação de micro-dados de pesquisas, tais como, a *Current Population Survey (CPS)* os procedimentos relacionados com a questão do sigilo são determinados pelo *Bureau of the Census*, que conduz essa pesquisa por um acordo entre agências.

Por outro lado, o Órgão Oficial de Estatística da Holanda, The *Netherlands Central Bureau of Statistics (CBS)*, tem uma política de disseminação de dados estatísticos, tanto para dados tabulados como para micro-dados, bastante organizada com relação à questão do sigilo, como pode ser vista nos artigos de Bethlehem & Keller(1990), Willenborg et al(1990) e Willenborg(1993).

Com relação à disseminação de micro-dados, o CBS somente divulga dados de pesquisas domiciliares, os da área econômica são proibidos por lei. São diferenciados dois tipos de micro-dados para usuários externos: para pesquisadores acadêmicos e para o público em geral. No primeiro caso, pela necessidade de riqueza de detalhes, os procedimentos para a limitação dos riscos de revelação não são tão rigorosos. Então, só são liberados os micro-dados para estatísticos empregados em instituições respeitáveis de pesquisa, mediante a assinatura de um contrato onde são especificados o propósito da pesquisa e as condições segundo as quais o dado pode ser usado, além do compromisso do usuário de não juntar os dados com os de outros arquivos, não fazer cópias a terceiros e submeter ao CBS os resultados obtidos a partir dos arquivos de micro-dados, para inspeção de possíveis riscos de revelação da informação.

São usadas técnicas para mascarar os dados, que consistem em remover características individuais raras, que permitam a identificação, A remoção é feita de alguma das seguintes maneiras: imputando valores ignorados para uma ou mais características nos registros, recodificando ou eliminando uma ou mais variáveis envolvidas, ou até mesmo eliminando os registros com os valores raros.

Para a liberação de micro-dados para o público em geral, os detalhes das variáveis de identificação são rigidamente restritos. Os mesmos procedimentos de mascarar os dados a serem utilizados por

acadêmicos são aplicados, porém as variáveis de identificação são reduzidas ao ponto de que nenhum identificador regional possa estar presente.

Com a introdução de regras para proteger a revelação para micro-dados foi desenvolvido, em 1991, um programa especial, denominado ARGUS. A intenção do CBS com o ARGUS é de torná-lo um programa para controle de revelação, não somente para liberação de micro-dados como também para tabulação. Não se trata de um programa feito sob medida, mas permite a especificação de outras regras, o que o torna útil para outros órgãos de estatística.

Para os dados tabulados a política do CBS é suprimir as células das tabelas, a partir da determinação de células potencialmente suscetíveis de identificação dos informantes.

Por fim, além dos artigos anteriormente mencionados, são ainda referenciados os artigos apresentados na *Annual Research Conference* de 1990, na Sessão sobre *Disclosure Avoidance*. *Veja Skinner(1990), Greenberg(1990) e Willenborg et al(1990).*

4. CONSIDERAÇÕES FINAIS

De uma análise superficial dos artigos mencionados, pode-se notar que a questão do sigilo é bastante complexa. Os avanços tecnológicos na área de computação oferecem oportunidades para processar, acessar e analisar grandes conjuntos de dados de forma mais eficiente, mas por outro lado facilitam o acesso a dados individualmente identificáveis e não autorizados a serem divulgados, além de reforçar a tendência dos usuários em solicitar dados cada vez mais detalhados.

A tarefa de encontrar o balanço entre o que é solicitado, acessível ou permitido não é de fácil execução.

A criação do grupo de trabalho para estudar as técnicas existentes, disseminá-las e propor as normas e procedimentos com vistas ao tratamento da questão do sigilo das informações no IBGE não é apenas recomendável, mas imperativa para que o IBGE desempenhe o seu papel de coordenador do SEN.

Diante da farta bibliografia a respeito do tratamento da questão do sigilo por órgãos internacionais, do artigo de Silva(1988) que trata da questão no IBGE (que adequadamente foi publicado na Série Textos para Discussão, mas que lamentavelmente ainda não foi discutido) e da necessidade emergente de criar normas e procedimentos para o tratamento da questão no IBGE, não há dúvidas de que os frutos do referido grupo de trabalho propiciarão o aumento do nível de credibilidade da instituição, da cooperação dos informantes e da satisfação dos usuários com os resultados acessíveis. Naturalmente, outro resultado desse grupo de trabalho será contribuir para o cumprimento do papel do IBGE, enquanto coordenador do Sistema Estatístico Nacional.

5 REFERÊNCIAS

BETHLEHEM, J.G.; Keller, W.J. e Pannekoek, J. Diclosure control of microdata. *Journal of the American Statistical Association*, 85, p. 38-45, 1990.

GREENBERG, B. Disclosure avoidance research at the Bureau of the Census. *Proceedings of Sixth annual Research Conference*. Washington, D.C.: Bureau of the Census, p. 144-66, 1990.

JABINE, T.B. Statistical disclosure limitation practices of United States Statistical Agencies. *Journal of Official Statistics*, v. 9, n. 2, p. 427-54, 1993.

SILVA, P.L.N. *O sigilo das informações estatísticas: idéias para reflexão*. Rio de Janeiro: IBGE, 1990. 49p. (Textos para Discussão, n. 4).

SKINNER , C.J. Disclosure avoidance for census microdata in Great Britain. *Proceedings of the Sixth Annual Research Conference*. Washington, D.C.: U.S. Bureau of the Census, p. 131-43, 1990.

WILLENBORG, L.C.R.J; Mokken, R.J. e Pannekoek, J. Microdata and disclosure risks. *Proceedings of the Sixth Annual Research Conference*. Washington, D.C.: U. S. Bureau of the Census, p. 167- 80, 1990.

WILLENBORG, L.C.R.J. Discussion: Statistical disclosure limitation. *Journal of Official Statistics*, v. 9, n. 2, p.469- 74, 1993.

ANEXO

JOS
JOURNAL OF OFFICIAL STATISTICS
Vol. 9, No. 2, 1993
© Statistics Sweden
ISSN 0282-423X

Contents

Privacy and Advances in Social and Policy Sciences: Balancing Present Costs and Future Gains

Paul D. Reynolds[1]

Abstract: Individuals have a right to an efficient, effective, and just society. Social science and policy research has provided considerable information about and understanding of some of the most critical issues in modern societies. This has often required extensive, sensitive data on individuals and organizations. As research progresses, continued scientific progress will require more complete and diverse data to further advance understanding of critical societal issues: health, crime, education, economic growth, etc. Increasing concerns regarding individual privacy and organizational confidentiality may lead to restrictions on access to sensitive, critical information on individuals and organizations as well as the capacity to assemble information from diverse public data sets. These dilemmas will be a continuing issue: only temporary compromises can be expected. It is proposed that the optimal strategy for the social science community is multi-faceted: clearly demonstrate respect for individuals and organizations through appropriate handling of private and sensitive information while simultaneously making clear the societal benefits of an enhanced understanding of basic phenomena.

Key words: Ethics; confidentiality.

Journal of Official Statistics
Vol. 9 No. 2, 1993, pp. 313–331
Statistics Sweden

Measures of Disclosure Risk and Harm

Diane Lambert[1]

Abstract: Disclosure is a difficult topic. Even the definition of disclosure depends on the context. Sometimes it is enough to violate anonymity. Sometimes sensitive information has to be revealed. Sometimes a disclosure is said to occur even though the information revealed is incorrect. This paper tries to untangle disclosure issues by differentiating between linking a respondent to a record and learning sensitive information from the linking. The extent to which a released record can be linked to a respondent determines disclosure risk: the information revealed when a respondent is linked to a released record determines disclosure harm. There can be harm even if the wrong record is identified or an incorrect sensitive value inferred. In this paper, measures of disclosure risk and harm that reflect what is learned about a respondent are studied, and some implications for data release policies are given.

Key words: Anonymity; confidentiality; disclosure threat; identification; linking, masked data; security.

Journal of Official Statistics
Vol. 9, No. 2, 1993, pp. 345–360
© Statistics Sweden

Informed Consent in U.S. Government Surveys

Robert H. Mugge[1]

Abstract: Informed consent is discussed as it is applied in U.S. Government social surveys and other information requests. From a survey of government surveys a representative group of 16 data programs is reviewed for the adequacy and propriety of their notifications to respondents. Special attention is given to requests for Social Security Numbers. The notifications are generally found to meet legal and moral requirements. Various implications of the notifications and major issues are discussed.

Key words: Privacy; confidentiality; social surveys.

Journal of Official Statistics
Vol. 9, No. 2, 1993, pp. 361–375
© Statistics Sweden

Informed Consent and Survey Response: A Summary of the Empirical Literature

Eleanor Singer[1]

Abstract: This paper reviews the published literature on the consequences of informed consent procedures for the conduct of social research. It examines empirical studies of four elements of consent – information concerning the content of the interview and the purposes of the research; assurances of confidentiality or anonymity; active versus passive consent; and information concerning voluntary participation – asking in each case what the effect of the factor is on response rate, response quality, and respondent reactions. Because much of the research is more than ten years old and because issues of privacy and confidentiality appear to be more salient to respondents than ever before, a program of research into these issues would seem to be a useful undertaking.

Key words: Informed consent; social science; confidentiality; voluntary participation; passive consent.

Journal of Official Statistics
Vol. 9, No. 2, 1993, pp. 383–406
Statistics Sweden

Masking Procedures for Microdata Disclosure Limitation

Wayne A. Fuller[1]

Abstract: Masking methods in which error is added to the elements of a data set prior to data release are described. The nature of confidentiality protection provided by these methods, the costs of such procedures to data providers, and the costs of the procedures to data users are investigated. Estimation methods appropriate for data subjected to a random mask are described. Nonlinear functions of the data vector are included in the estimation discussion.

Key words: Confidentiality; measurement error; data transformations.

Journal of Official Statistics
Vol. 9, No. 2, 1993, pp. 407–426
© Statistics Sweden

Statistical Analysis of Masked Data

Roderick J.A. Little[1]

Abstract: A model-based likelihood theory is presented for the analysis of data masked for confidentiality purposes. The theory builds on frameworks for missing data and treatment assignment, and a theory for coarsened data. It distinguishes a model for the masking selection mechanism, which determines which data values are masked, and the masking treatment mechanism, which specifies how the masking is carried out. The framework is applied to a variety of masking methods, including randomized response, subsampling of cases or variables, deletion, coarsening by grouping or rounding, imputation, aggregation, noise injection and simulation of artificial records.

Key words: Confidentiality; grouped data; imputation; missing data; randomized response; rounded data; slicing.

Journal of Official Statistics
Vol. 9, No. 2, 1993, pp. 427–454
© Statistics Sweden

Statistical Disclosure Limitation Practices of United States Statistical Agencies[1]

Thomas B. Jabine[2]

Abstract: One of the topics examined by the Panel on Confidentiality and Data Access was the use of statistical disclosure limitation procedures to limit the risk of disclosure of individual information when data are released by U.S. federal statistical agencies in tabular or microdata formats. To assist the Panel in its review, the author prepared a summary of the disclosure limitation procedures that were being used by the agencies early in 1991. This paper is an updated version of that summary.

Key words: Statistical disclosure limitation; tabulations; microdata.

Journal of Official Statistics
Vol. 9, No. 2, 1993, pp. 475–499
© Statistics Sweden

Database Systems: Inferential Security

Sallie Keller-McNulty[1] *and Elizabeth A. Unger*[2]

Abstract: The problems of data security and confidentiality have been studied by computer scientists and statisticians. The areas of emphasis within these disciplines on data security are different but not disjoint. One of the main differences is how one views data release. Statisticians have focused on aggregate data release and on single static files of microdata records. Computer scientists have focused on data release through sequential queries to a database. An initial integrating factor of the two fields is the concept of information stored as a federated database. This paper synthesizes the research done in both of these disciplines and provides an extensive review of the literature. Some basic definitions integrating the two fields are given and data security and confidentiality methodologies studied in both disciplines is discussed.

Key words: Data security; database security; disclosure; compromise.