

Aperfeiçoamentos no processo de acesso a microdados restritos no IBGE¹

Maria Luiza Barcellos Zacharias (maria.zacharias@ibge.gov.br)
Zélia Magalhães Bianchini (zelia.bianchini@ibge.gov.br)
Sonia Albieri (sonia.albieri@ibge.gov.br)

Resumo

O notável desenvolvimento da Tecnologia da Informação nos últimos anos alterou significativamente a maneira como os Institutos Nacionais de Estatística (INE) lidam com seus usuários. De um lado, os INE têm buscado atender as diferentes necessidades de usuários para acesso a dados, especialmente as demandas provenientes de pesquisadores que exigem informações muito detalhadas e bancos de dados precisos para o seu trabalho. Por outro lado, os Institutos Nacionais de Estatística têm a obrigação legal e moral de respeitar a privacidade dos entrevistados. Como consequência, vários INE têm desenvolvido procedimentos e investido em soluções a fim de encontrar o equilíbrio certo para enfrentar esse desafio.

Seguindo essa tendência, desde 2003 o IBGE tem disponibilizado aos pesquisadores o acesso restrito a arquivos de dados não disponíveis para o público em geral, mediante o uso de uma sala especialmente destinada para este objetivo, na sede do instituto, onde os pesquisadores podem, então, realizar análises mais detalhadas e aprofundadas.

O objetivo deste artigo é descrever os recentes aprimoramentos sobre os procedimentos que têm sido adotados pelo IBGE a fim de minimizar o risco de revelação de informantes e aperfeiçoar o processo de acesso a dados restritos por pesquisadores, bem como para dar mais transparência para a sociedade, fornecendo informações sobre os métodos disponíveis para acessar os dados, mantendo a proteção à confidencialidade dos dados de pesquisas e de censos.

1. Introdução

1. Tradicionalmente, a disseminação de estatísticas no IBGE vem sendo realizada de duas formas: para o público em geral, através dos meios de comunicação, por meio de divulgação de resultados (*press releases* preparados pela área técnica em conjunto com a área de comunicação social) ou em conferências de imprensa; e para os usuários em geral, por meio de publicações impressas e em meio eletrônico. Quando os usuários são mais especializados ou pertencem a órgãos de governo, as demandas geralmente são atendidas por meio de tabulações especiais ou de arquivos de microdados de uso público.

2. Desde 2001, o IBGE estabeleceu uma política de disseminação gratuita de todos os seus

¹ Tradução livre do artigo *Improvements on risk management for micro-data access at the Brazilian Institute of Geography and Statistics* escrito para a *Conference of European Statisticians*, realizada em Genebra, Suíça, de 10 a 12 de junho de 2013.

Disponível em <http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2013/40.pdf>

produtos na Internet, o que gerou um significativo aumento do volume de acesso a esse canal de comunicação. Desde então, o número de páginas de todas as publicações em papel vem diminuindo, uma vez que as publicações têm um CD-ROM encartado contendo muitas tabelas adicionais além das apresentadas nos volumes impressos.

3. Além das publicações eletrônicas na página do IBGE na internet, vale destacar duas bases de dados importantes disponíveis para acesso: o SIDRA - Sistema de Recuperação de Dados Agregados - banco de dados com informações agrupadas em nível territorial que permite aos usuários construir tabelas de acordo com informações selecionadas do banco de dados - e o BME - Banco Multidimensional de Estatísticas - banco de dados com informações de microdados que permitem aos usuários construir tabelas de acordo com informações selecionadas e sujeitas às restrições de confidencialidade. Esse último requer que o usuário faça uma assinatura para acesso ao serviço.

4. Além disso, desde o início da década de 1990, o IBGE divulga arquivos de microdados de uso público das pesquisas domiciliares por amostragem probabilística. Para proteger a confidencialidade desses microdados, adota-se a supressão de dados para níveis geográficos detalhados. Por outro lado, não há liberação de arquivos de microdados das pesquisas econômicas e do Censo Agropecuário, nem do questionário básico do Censo Demográfico, que investiga informações para toda a população.

5. Apesar da ampla divulgação dos dados do IBGE por meio de publicações impressas e na internet, a impossibilidade de liberar os arquivos com dados mais detalhados limita a utilidade desses dados para fins de pesquisa e de políticas públicas.

6. A crescente demanda por microdados detalhados, o avanço da tecnologia e o aumento da preocupação com questões de privacidade levaram o IBGE, a partir de 2003, a propiciar acesso pelos pesquisadores a arquivos de dados que não são liberados para o público em geral, permitindo análises mais aprofundadas do que aquelas possíveis com dados tabulados ou agregados. Em resposta ao interesse da comunidade de pesquisa, o IBGE criou uma sala de acesso especial a dados restritos onde os pesquisadores podem acessar arquivos de dados detalhados em um ambiente seguro, sem comprometer a confidencialidade dos entrevistados.

7. Ao longo dos últimos 10 anos, cerca de 200 projetos de pesquisa foram submetidos e analisados pelo Comitê de Avaliação de Acesso a Microdados Não Desidentificados. Destes 200, 73% referem-se a dados de pesquisas de empresas, a maioria relacionados à Pesquisa de Inovação Tecnológica. Outros pedidos referem-se ao Censo Agropecuário, ao Censo Demográfico e às Pesquisas Domiciliares.

8. Recentemente, o IBGE reconheceu a necessidade de rever algumas diretrizes, regras e procedimentos que foram adotados para os pesquisadores terem acesso aos arquivos de dados restritos. Para tanto, foi realizado um amplo estudo sobre os procedimentos e práticas existentes e desenvolvidos por outros institutos nacionais de estatística, para servir de base para as melhorias que serão implementadas no segundo semestre de 2013. Esta revisão das melhores práticas provocou alterações nos procedimentos de uso da sala de acesso de dados restritos, conforme descrito abaixo.

2. Situação atual e aperfeiçoamentos

9. Os atuais procedimentos para obtenção de acesso a arquivos de microdados para fins científicos estão descritos a seguir, juntamente com as mudanças e melhorias que estão sendo implementadas.

(a) Submissão do projeto de pesquisa

Para poder acessar dados restritos na sala destinada a este fim, os pesquisadores devem, inicialmente, apresentar uma proposta de projeto de pesquisa, em estilo livre, obedecendo as instruções a respeito das informações que devem ser incluídas no pedido. Entretanto, algumas vezes os pedidos são devolvidos ao pesquisador para completar informações faltantes. Para resolver este problema, um formulário padronizado está sendo desenvolvido, contendo campos específicos a serem preenchidos com as informações solicitadas para serem analisadas para a concessão do acesso. Este formulário estará disponível na página do IBGE na internet, permitindo que a proposta seja submetida eletronicamente, e não mais em papel. Espera-se que esta mudança reduza tanto o número de devoluções de propostas para complementação quanto o intervalo de tempo entre a solicitação de dados e a aprovação do pedido.

Além disso, um guia do usuário está sendo elaborado e estará disponível para download na página do IBGE na internet, para que os pesquisadores fiquem cientes de todos os procedimentos e regras a serem seguidos para obter autorização de uso da sala de acesso a dados restritos. Esta iniciativa visa garantir a transparência para o público, a fim de evitar questionamentos sobre eventual violação de confidencialidade. O IBGE considera uma boa prática informar que um dos usos dos dados das pesquisas será o acesso a microdados desidentificados por pesquisadores, sob condições controladas e para fins específicos.

(b) Avaliação do projeto

O Comitê de Avaliação de Acesso a Dados Restritos é responsável por avaliar o projeto, com base em informações fornecidas pela área temática responsável pela produção dos dados da pesquisa. O Comitê autoriza (ou não) o uso da sala de acesso a dados restritos, assegurando que não haverá risco de quebra de confidencialidade.

O Comitê é presidido pela Diretora de Pesquisas Substituta e composto por membros seniores das equipes das áreas de pesquisas de empresas, de metodologia e de disseminação de dados.

Com o objetivo de aperfeiçoar esta etapa, está sendo desenvolvido um formulário padrão para avaliar as propostas, para que sejam realizadas avaliações padronizadas entre as diferentes áreas temáticas. Além disso, um sistema eletrônico para gerenciar todo o processo está em desenvolvimento para substituir o atual sistema de controle dos projetos submetidos, que ainda é feito manualmente.

A incorporação de bases de dados externas aos arquivos de microdados solicitados geralmente é permitida, a menos que se perceba a existência de algum risco de revelação da identidade dos informantes. Atualmente, essa avaliação é feita pelo analista temático de forma não estruturada, antes da junção dos dados. A ideia é que a avaliação do resultado deste procedimento seja mais rigorosa de modo que o risco de quebra de confidencialidade seja reduzido.

Outra iniciativa para tornar esta incorporação de dados externos mais restritiva é exigir que os pesquisadores apresentem uma autorização formal, assinada pelos gestores das bases de dados externas, para garantir que houve consentimento da fonte externa para utilização daqueles dados no projeto de pesquisa.

(c) Termos de compromisso para acesso aos dados

Depois do projeto de pesquisa ter sido aprovado, um termo de compromisso é assinado entre o pesquisador e o instituto de estatística, especificando as condições e os valores das taxas de uso da sala de acesso. Está sendo estudada a imposição da necessidade de os pesquisadores serem associados a uma universidade ou a um instituto de pesquisa para terem autorização de uso da sala de acesso. Neste caso, as autoridades

dessas entidades também deverão assinar o termo de compromisso e serem corresponsáveis pelas regras de acesso e de utilização dos dados. Atualmente, não há nenhuma restrição para um pesquisador independente ter acesso aos microdados na sala de acesso.

(d) Acesso local

Na sala de acesso, os arquivos com os microdados restritos são instalados em computadores especiais munidos de aspectos de segurança, tais como bloqueio à rede e à internet para transferência de dados, e desativação das unidades de disco externas e das portas paralelas e seriais. Os bancos de dados de pesquisas de empresas do IBGE ou de fontes externas têm a identificação das empresas criptografada.

Após a conclusão do trabalho, o pesquisador grava as saídas produzidas no disco rígido do computador e, em seguida, elabora um relatório documentando o que foi feito. Estas informações são então gravadas em um CD-ROM para serem posteriormente analisadas por especialistas da área temática responsável pela base de dados da pesquisa. Para aprimorar este procedimento, está em desenvolvimento o uso de uma rede interna de dados protegida para transmitir os arquivos de dados, para reduzir o risco de quebra de confidencialidade, no caso de perda do CD-ROM.

Outra melhoria a ser implementada é que apenas as variáveis solicitadas, e que sejam realmente necessárias para a realização das análises propostas, serão incluídas no arquivo de microdados, de acordo com o princípio “need to know”. Hoje, todas as variáveis da pesquisa são incluídas no arquivo para o pesquisador, mesmo que não tenham sido solicitadas.

(e) Verificação e liberação das saídas

As saídas produzidas pelo pesquisador devem ser analisadas antes de serem liberadas, de modo a garantir a avaliação técnica do risco de identificação dos informantes e dos requisitos de confidencialidade. A verificação das saídas é feita pela área temática responsável pelos microdados da pesquisa. Depois de ser aprovada sua liberação, ou seja, depois que a área temática avalia que não há risco de revelação, outro termo de compromisso é assinado.

Este termo estabelece as condições de uso dos dados gerados por esta modalidade especial de acesso, onde o usuário reconhece que os dados são de propriedade do IBGE, e se compromete a informar que os resultados e análises eventualmente divulgados envolvendo esses dados foram obtidos por meio desse acesso especial.

Para aprimorar a avaliação técnica do risco de revelação, um grupo técnico composto por estatísticos vêm realizando estudos sobre os métodos estatísticos relacionados à confidencialidade, para implementação futura em todos os arquivos de dados e tabelas produzidas pelo IBGE, inclusive nos arquivos de dados restritos utilizados pelos pesquisadores. Atualmente, o processo de anonimização (desidentificação) dos dados consiste apenas na remoção das variáveis de identificação e na agregação ou supressão de células para evitar o risco de revelação da identidade dos informantes.

3. **Trabalhos futuros**

10. O Princípio 4 do documento "Diretrizes e princípios fundamentais para a gestão do sigilo

estatístico e do acesso a microdados”² recomenda: "Os processos de acesso a microdados por pesquisadores, bem como os usos e os usuários de microdados, devem ser transparentes e disponíveis ao público"

11. A página do IBGE na Internet é um meio efetivo para fornecer informações sobre como os pesquisadores podem ter acesso a microdados restritos. Atualmente, os procedimentos para submissão de projetos de pesquisa estão disponíveis apenas na intranet, sendo que os usuários são informados sobre esses procedimentos apenas quando solicitam acesso a dados especiais.

12. Em virtude disto, o IBGE passará a dar publicidade e transparência sobre o acesso a dados restritos, destacando que os dados serão fornecidos aos pesquisadores sob condições controladas e para fins específicos de pesquisa. A visibilidade destes procedimentos é necessária para aumentar a confiança pública de que os microdados serão usados de forma segura e adequada. O objetivo é ser totalmente transparente sobre os usos específicos de microdados para evitar suspeitas de mau uso e para assegurar que os pesquisadores estejam cientes das consequências, para eles próprios e para a sua instituição, caso haja violação das regras estabelecidas.

4. Conclusão

13. É fundamental buscar atender às necessidades da comunidade de pesquisadores, garantindo a confidencialidade dos dados e evitando o risco de revelação da identidade dos informantes. O acesso a dados restritos para fins de pesquisa requer a colaboração de todas as pessoas envolvidas, ao mesmo tempo em que o Instituto de Estatística deve estar preparado para lidar com uma série de situações desafiadoras. Apesar das limitações, o IBGE está se preparando para enfrentar tais desafios, aprimorando os procedimentos de acesso a dados restritos, visando atender às necessidades dos pesquisadores, ao mesmo tempo em que garante a confidencialidade dos informantes.

14. Neste sentido, tendo em mente as diretrizes e princípios fundamentais dos institutos de estatística, o IBGE vem buscando manter-se atualizado com a experiência de outros institutos nacionais de estatística e com o estudo de métodos estatísticos de confidencialidade, saindo de uma condição de evitar o risco de revelação para uma condição de gerenciar este risco, no processo de fornecimento de acesso a microdados restritos.

5. Referências

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 Microdata Access. Principles and guidelines of good practice – interim guidelines. Submitted by the Task Force on Managing Confidentiality and Microdata 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.

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

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