IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA
DPE - DIRETORIA DE PESQUISAS
NME - NÚCLEO DE METODOLOGIA

O SIGILO DAS INFORMAÇÕES ESTATÍSTICAS

IDÉIAS PARA REFLEXÃO

TEXTOS PARA DISCUSSÃO

-VOLUME I NÚMERO 4

ABRIL DE 1988 DPE-88004

Pedro Luis do Nascimento Silva

APRESENTAÇÃO

Este texto foi elaborado com a finalidade precípua de subsidiar o debate interno na Diretoria de Pesquisas; durante. as atividades preparatórias do Seminário de Disseminação de Informações do IBGE realizado em dezembro de 1987 e da III CONFEST - Conferência Nacional de Estatística, a ser realizada em 1988. São abordados diversos aspectos do problema: legal, ético, político e técnico/operacional.

O texto baseia-se no conhecimento superficial dos procedimentos e métodos empregados para a manutenção do sigilo nas diversas pesquisas e levantamentos realizados pelo IBGE, e no artigo “On Disclosure Control of Micro Data”, de Keller, W.J. Bethlehem, J.G. (1987).

Apresenta-se como principal recomendação 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.

1. INTRODUÇÃO

O problema da manutenção do sigilo das informações estatísticas precisa ser abordado segundo seus diversos aspectos relevantes, que compreendem um lado legal, um lado político, um lado ético e uma face técnico-operacional. Ignorar qualquer desses aspectos é perigoso, e pode conduzir a decisões cujo efeito será desastroso para a credibilidade da organização.

Tendo isso em mente, e identificando no momento atual do IBGE a oportunidade para se discutir o problema em profundidade, visando à formulação de uma política de disseminação de informações, procura-se abordar nas seções seguintes cada um dos diferentes aspectos do problema já citados.

É oportuno também registrar que, apesar de não se dar de forma organizada, já existe um debate interno na instituição sobre o assunto, motivado pela crescente pressão dos usuários externos na tentativa de obter acesso aos microdados das pesquisas do IBGE, ou mesmo a dados agregados: para níveis geográficos bastante detalhados, como o setor censitário. Neste debate, as áreas técnicas da instituição responsáveis pelo atendimento a esses usuários externos tendem a sensibilizar-se pelas pressões, e desempenham papel questionador junto às áreas produtoras que, em geral, são menos vulneráveis à pressão, além de dispensarem atenção menor ao problema.

2. O PROBLEMA LEGAL

A legislação referente ao sigilo de informações é a seguinte:

- **Lei n° 5534, de 14/11/68** - dispõe sobre a obrigatoriedade de prestação de informações estatísticas e dá outras providências - esta é a Lei básica sobre o assunto, e trata da obrigatoriedade de prestação de informações solicitadas pelo IBGE para execução dos levantamentos e pesquisas constantes do plano Nacional de Estatística baixado pelo **decreto-lei n°161, de 13/02/67**, impondo como contrapartida que "as informações prestadas terão caráter sigiloso, serão usadas exclusivamente para fins estatísticos, e não poderão ser objeto de certidão nem... servirão de prova em processo administrativo, fiscal ou judicial...". Esta Lei não estabelece qualquer dispositivo para punição dos infratores deste dispositivo que obriga a manutenção do sigilo das informações, preocupando-se quase que exclusivamente com o problema de obrigatoriedade de prestação de informações.

- **Lei n. 5878, de 11/05/73** - dispõe sobre a Fundação IBGE e dá outras providências - esta Lei estabelece a estrutura atual do IBGE, e modifica a anterior ao estabelecer no seu artigo 6° que “As informações necessárias ao Plano Geral de Informações Estatísticas e Geográficas serão prestadas obrigatoriamente pelas pessoas naturais e pelas pessoas jurídicas de direito público e privado e utilizadas exclusivamente para os fins a que se destinam, não podendo servir de instrumento para qualquer procedimento fiscal ou legal contra os informantes,...” e também no seu artigo 8° que “Para desempenho de suas atribuições, o IBGE poderá firmar acordos, convênios e contratos com entidades públicas e privadas, preservados o sigilo e o uso das informações e os interesses da segurança nacional”. Esta Lei parece apontar a possibilidade de cessão das informações para outras

entidades, quando estas estiverem realizando papéis de complementação das atividades do IBGE, desde que devidamente autorizadas por acordos, convênios ou contratos.

- **Decreto n° 73177, de 20/11/73** - este Decreto regulamenta a Lei n° 5534, de 14/11/68, modificada pela Lei n° 5878, de 11/05/73, e dispõe sobre a obrigatoriedade da prestação de informações necessárias ao Plano Nacional de Estatísticas Básicas e ao Plano Geral de Informações Estatísticas e Geográficas. Em termos de conteúdo, o Decreto nada acrescenta com respeito à questão do sigilo, detalhando apenas os procedimentos para punição dos responsáveis pela recusa de informações so1icitadas pelo IBGE.

- **Decreto n° 74084, de 20/05/74** - este Decreto aprova o Plano Geral de Informações Estatísticas e Geográficas, e no seu artigo 8° faz referência às Leis anteriores sobre o sigilo, estabelecendo que "As informações resultantes dos levantamentos previstos no Plano só poderão ter a utilização referida no artigo 6° da Lei n° 5878, estando protegidas pelo sigilo assegurado pelo artigo 1°, parágrafo único, da Lei n° 5534". No seu artigo 9°, dispõe que "As informações resultantes dos levantamentos previstos no Plano..., depois de devidamente processadas pelos meios indicados, e atendidas, em cada caso, as normas e exigências sobre o assunto, serão divulgadas pelo IBGE e postas à disposição dos interessados, através de anuários, relatórios, sinopses, mapas, cartas topográficas, cartas temáticas, publicações especializadas e demais formas de divulgação".

- **Decreto n. 77624, de 17/05/76** - este Decreto "dispõe sobre a utilização, pelo IBGE, de dados informativos de origem governamental na produção de informações e estudos..." Ele disciplina o acesso pelo IBGE às "informações estatísticas existentes nos órgãos e entidades da administração civil, direta e indireta, e nas fundações supervisionadas". Ressalva, no parágrafo 1° do artigo l° que "nos casos em que houver sigilo a ser resguardado, tal circunstância será prévia e expressamente comunicada ao IBGE pelo órgão, entidade ou fundação fornecedor dos dados", e no parágrafo 2° do mesmo artigo que, nesta hipótese "o IBGE dará tratamento especial aos dados recebidos sendo o responsável pela rigorosa observância do disposto na legislação precedente sobre o sigilo".

Não é difícil perceber que a legislação estabelece uma relação de troca entre órgãos produtores de estatísticas (no caso, prioritariamente o IBGE) e os informantes dos levantamentos e pesquisas constantes do Plano Geral de Informações Estatísticas e Geográficas, no sentido de obrigar os primeiros a manter em sigilo as informações obrigatoriamente prestadas pelos últimos.

Por outro lado, a questão parece esgotada no âmbito das Leis e Decretos, não havendo normas técnicas ou operacionais disponíveis que detalhem e disciplinem os procedimentos que devem ser adotados no sentido de assegurar a manutenção do sigilo previsto em Lei. A inexistência desses instrumentos normativos e orientadores das ações operacionais tem possibilitado a existência de diferentes interpretações sobre o que vem a ser informações sigilosas, quem pode ter acesso a essas informações e que informações individuais podem ser divulgadas.

Na maioria dos casos, as interpretações se consolidaram em procedimentos tradicionais, com a chamada "desidentificação das tabulações das pesquisas econômicas" ou o fornecimento de cadastros em listagens ou meios magnéticos, onde os dados de identificação dos informantes são incluídos e dados substantivos recolhidos das pesquisas não; no caso dos

cadastros retirados dos Censos Agropecuários, a prática é diferente, e o fornecimento das informações depende da natureza do usuário solicitante e do uso que este pretende dar aos dados recebidos.

Todas essas diferenças de comportamentos podem ser explicadas pela inexistência de normas claras e objetivas que definam uma interpretação oficial da legislação existente, que passe a ser adotada por todos os departamentos e coordenadorias responsáveis por censos e que discipline o processo de disseminação de informações, orientando também o relacionamento das áreas produtoras com as áreas de disseminação.

A necessidade de criar grupo de trabalho envolvendo ambas as áreas, encarregado de propor a normatização do tratamento do sigilo das informações estatísticas, servirão de motivação para as recomendações apresentadas na última seção deste documento.

3. O PROBLEMA ÉTICO

O lado ético do problema do sigilo das informações estatísticas não pode ser ignorado. Tudo começa no próprio momento em que um agente credenciado pela Instituição de pesquisa aborda o informante para coletar informações. Neste mesmo instante, é assumido pelo órgão que coleta as informações um compromisso tácito junto aos seus informantes no sentido da manutenção do sigilo das informações individualizadas.

Muitas vezes, na tentativa de obter a colaboração do informante, o agente faz referência à legislação que garante a obrigatoriedade da prestação das informações solicitadas e assegura, em contrapartida, o sigilo no tratamento das informações prestadas.

No IBGE, até mesmo alguns instrumentos de coleta fazem referência à legislação que garante o sigilo das informações estatísticas, na tentativa de motivar a colaboração dos informantes. Um exemplo disso são as instruções de preenchimentos dos questionários do Censo Econômico de 1985, onde se registra que "a legislação dos Censos Econômicos de 1985 mantém a caráter obrigatório e confidencial atribuído às informações coletadas pelo IBGE, as quais se destinam, exclusivamente, a fins estatísticos e não poderão ser objeto de certidão e nem terão eficácia jurídica como meio de Prova".

Além disso, o Manual do Recenseador também registra que "em hipótese alguma os questionários e informações podem ser vistos por pessoas estranhas ao trabalho censitário", o que configura falta passível de punição.

Por outro lado, a "Declaração Sobre Ética Profissional" do Instituto Internacional de Estatística (ISI) do qual o IBGE é membro ex-ofício, estabelece que "os Estatísticos deverão adotar medidas apropriadas para impedir que seus dados sejam publicados ou divulgados por qualquer outro meio ou forma que possibilite a descoberta ou inferência da identidade de qualquer indivíduo".

Além disso, o Código de Ética dos Estatísticos profissionais no Brasil, baixado pela Resolução número 58 de 06/10/76 do Conselho Federal de Estatística, estabelece que "no exercício de suas funções, é dever precípuo do Estatístico guardar sigilo dos assuntos que lhe chegarem ao conhecimento em razão de seus deveres profissionais". Embora não seja muito

enfático, e não se refira diretamente à manutenção do sigilo das informações, deixa implícita a preocupação com o trato sigiloso dos assuntos que chegarem ao conhecimento do profissional em razão da natureza de sua atividade.

De qualquer modo, verifica-se que a essência do problema pela ética é a importância da colaboração dos informantes no sentido da prestação de informações fidedignas aos órgãos produtores de estatísticas. Sem essa colaboração, torna-se praticamente inviável a montagem de um sistema Estatístico capaz de fornecer informações precisas, confiáveis e pontuais.

A implantação de códigos de ética como os citados é uma tentativa de evitar que, pela atuação indevida de um profissional, todo o sistema Estatístico seja prejudicado pela recusa dos informantes em colaborar. Está implícita nesta idéia a hipótese de que os informantes deixarão de colaborar caso não sintam segurança quanto ao sigilo das informações que prestaram aos órgãos de estatística.

É oportuno registrar que, apesar, da plena vigência dos códigos de ética citados, estes são desconhecidos em parte ou no todo pela maioria dos Estatísticos, e totalmente ignorados pelos demais profissionais envolvidos na coleta, tratamento, guarda e disseminação das informações estatísticas. Os principais interessados na aplicação zelosa dos princípios éticos ali contidos, os órgãos que dependem da colaboração dos informantes para justificarem sua própria existência, como o IBGE, pouco ou nada fazem no sentido de divulgar entre seus funcionários esses códigos de ética, motivo pelo qual fica até difícil exigir mais tarde um comportamento plenamente compatível com os princípios ali estabelecidos.

A esse respeito, a principal iniciativa que se poderia recomendar é a ampla disseminação dos códigos de ética entre os funcionários mediante treinamento, e a fixação de normas de conduta próprias para serem seguidas pelos funcionários da organização.

Isto poderia contribuir de maneira significativa para um tratamento mais adequado do problema da manutenção do sigilo das informações estatísticas, tanto pelos milhares de pessoas envolvidas na coleta e apuração das pesquisas, como pelos responsáveis pela guarda e disseminação das informações produzidas. Fica evidente a necessidade de envolver na tarefa de preservação do sigilo das informações praticamente todos os funcionários da instituição, ao invés de deixá-la sob responsabilidade exclusiva dos Estatísticos.

4. O PROBLEMA POLÍTICO

Um dos pontos - chave da questão do sigilo é de natureza essencialmente política: é a credibilidade da organização. Qualquer instituição produtora de informações estatísticas se apoia fundamentalmente na sua credibilidade. A aceitação pelo público das informações produzidas, a colaboração do público na prestação de informações, e mesmo a obtenção de recursos para realização de pesquisas dependem da credibilidade da organização.

Mas a manutenção do sigilo das informações é um dos fatores que influenciam de forma significativa a credibilidade da organização junto aos seus informantes e usuários. A ocorrência de violação do sigilo das informações pode destruir, de um momento para o outro, a credibilidade de uma organização, muitas vezes conquistada através de anos de trabalho

sério e árduo. Demonstra-se assim, uma vez mais, a importância do cuidado com a preservação do sigilo.

Outro aspecto importante, do ponto de vista político, é a crescente preocupação das pessoas com a privacidade. Em alguns países, esse fato tem preocupado bastante os órgãos produtores de estatística, pois há inclusive segmentos expressivos e organizados da sociedade empenhados em campanhas destinadas a impedir o controle da sociedade pelo Estado, o que significa, por exemplo, a recusa em prestar informações aos órgãos do governo.

Na Alemanha, uma campanha desse tipo motivou recentemente o cancelamento da realização de um Censo Demográfico. Na Holanda, as pesquisas domiciliares atingiram um nível de recusa/não resposta da ordem de 35%, o que demonstra a dificuldade que o órgão produtor de estatística tem de atuar num cenário como esse. Nesses países, somente através de garantias reais do sigilo das informações estatísticas se consegue ainda obter a colaboração de parte da população.

No Brasil, o panorama ainda é bastante diferente: há uma legislação impondo a obrigatoriedade da prestação de informações para fins estatísticos, e a preocupação com a privacidade ainda é secundária, ou pelo menos, a privacidade não preocupa a grande maioria da população.

No entanto, há uma crise grave de credibilidade das instituições oficiais, em particular do IBGE, principalmente devido ao mau uso pelo governo das estatísticas produzidas. Num cenário como esse, o problema do sigilo pode parecer secundário.

Por outro lado, a ocorrência de incidentes relacionados com a violação do sigilo de informações poderia aumentar substantivamente a crise de credibilidade, e afetar de maneira significativa a cooperatividade dos informantes.

Levando tudo isso em conta, e avaliando as perspectivas futuras de evolução do quadro político, parece recomendável a adoção de providências que assegurem a manutenção do sigilo das informações, de modo a construir uma imagem de credibilidade para a instituição, que inspire a confiança e cooperação dos informantes.

Sem essa imagem de credibilidade, o crescimento da conscientização da população sobre a privacidade pode levar a organização a situações extremas, onde os informantes se recusem a prestar as informações solicitadas ou a colaborar de qualquer maneira

5. O PROBLEMA TÉCNICO-OPERACIONAL

É neste aspecto que a situação parece mais difíci1. Ao longo dos anos o IBGE consolidou, em suas diversas áreas operacionais responsáveis pela produção e disseminação de pesquisas, o uso de práticas tradicionais destinadas à manutenção do sigilo das informações que hoje parecem insuficientes e inadequadas, além de bastante heterogêneas entre as áreas.

O problema se agravou com o avanço do uso da informática tanto na produção como na disseminação de dados, que não foi acompanhado no IBGE por qualquer esforço técnico de

incorporação da nova realidade dentro dos procedimentos adotados para manutenção do sigilo das informações.

A1guns exemplos podem ser citados, sempre com a intenção de ilustrar o problema e não de criticar o trabalho das áreas envolvidas.

O primeiro exemplo digno de nota é o trabalho de "desidentificação", que é executado na fase preparatória da divulgação dos dados das pesquisas da área econômica (Censos Econômicos, Pesquisa Industrial Anual, Empresas de Transporte Rodoviário, Meios de Hospedagem) com o objetivo de impedir a individualização dos resultados a nível de informante. Esse trabalho está baseado na interpretação da legislação de que divulgar uma tabela onde se possa conhecer os dados referentes a 1 ou 2 informantes, mesmo sem que esses informantes sejam identificados na tabela, constitui uma violação do sigilo das informações.

Dai, quando houver células numa tabela qualquer onde ocorrem somente 1 ou 2 informantes, procede-se à desidentificação da célula, o que consiste em mascarar ou omitir com um símbolo (X) todos os resultados correspondentes a essa célula, à exceção do número de informantes.

Acontece que os resultados das pesquisas são geralmente apresentados segundo classificações, e nunca relacionados com dados de identificação propriamente ditos dos informantes, como endereço, nome ou razão social da firma, CGC, e outros.

A aplicação desse procedimento, como é concebido hoje em dia, exige o dispêndio de tempo e a utilização de mão-de-obra especializada na tarefa, sem contudo garantir de forma efetiva que não possa ocorrer individualização de dados.

Fica evidente que os procedimentos para proteção do sigilo dos dados adotados são essencialmente voltados para a divu1gação de dados através de publicações, não incorporando adequadamente o tratamento para outras formas de divulgação já em uso no IBGE, tais como sistemas de acesso direto à Base de Dados via telex ou terminal, arquivos em fitas magnéticas para "pronta entrega", e mesmo as tabulações especiais.

Outro exemplo de problema não tratado pelos atuais procedimentos de manutenção do sigilo das informações é causado pela descentralização do processo de apuração das pesquisas. Esse processo obriga que os dados sejam transmitidos das unidades regionais onde os questionários foram transcritos para meio magnético, para o órgão central por meio de ligações remotas entre os computadores. É sabido que essas ligações podem ser violadas com relativa facilidade, como comprovam as experiências de outros países de onde a tecnologia foi importada, caso não sejam adotadas precauções adequadas.

Os órgãos centrais de estatística da Austrália e da Suécia já estão operando toda a transmissão eletrônica de dados de forma criptografada, isto é, usando códigos cuja decifração é bastante difícil. Esta preocupação simples é capaz de praticamente eliminar os riscos de·violação do sigilo das informações durante a transmissão desses dados entre diferentes unidades daqueles órgãos.

Há outros aspectos técnicos a abordar. Por exemplo, será que a idade do dado influencia a questão do sigilo? Será que liberar em 1987 dados tabulares referentes ao Censo

Demográfico de 1970 em forma tal que pudessem ocorrer células com 1 ou 2 informantes consiste em violação do sigilo das informações?

E quanto a dar acesso às informações a outros órgãos produtores de estatísticas? Até o momento, o IBGE tem tratado de forma semelhante tanto órgãos produtores de informações estatísticas quanto órgãos fiscais, no que diz respeito à liberação do acesso aos microdados. Será que este procedimento é correto? Não deveriam as instituições idôneas, membros do Sistema Estatístico Nacional, ter um acesso maior aos microdados que aquele propiciado a órgãos de outra natureza? E neste caso, não seria coerente que os órgãos seguissem normas de conduta idênticas às adotadas pelo IBGE para tratar a questão da manutenção do sigilo, responsabilizando-se formalmente pelo seu cumprimento nos convênios e contratos firmados?

Como já foi dito anteriormente, a responsabilidade pela manutenção do sigilo das informações começa no mesmo instante em que as informações são coletadas junto aos informantes das pesquisas e levantamentos. Que medidas ou precauções são hoje efetivamente adotadas para garantir a manutenção do sigilo até que as informações cheguem ao órgão central, encarregado pela apuração, tabulação, armazenamento e disseminação das informações? E que medidas seria necessário adotar com essa finalidade?

Mesmo durante o processo de apuração e tabulação, e posteriormente quando as informações já se encontrarem armazenadas, que medidas são adotadas para impedir ou dificultar o acesso indevido ou a violação do sigilo das informações? Hoje em dia, pode-se "levar para casa" uma fita contendo os dados cadastrais de identificação junto com os dados substantivos coletados na Pesquisa Industrial Anual do IBGE, sem que qualquer dos departamentos responsáveis por essa pesquisa fiquem sabendo disso. Para isso, é suficiente ter uma sigla e ter acesso à documentação·dos arquivos, que é de domínio público no IBGE. Esse é um nível de risco aceitável? Se não, que medidas seria necessário adotar para minimizar os riscos de violação do sigilo das informações?

E quanto ao processo de disseminação: hoje em dia são produzidos diversos arquivos contendo microdados (dados a nível dos informantes) para disseminação dos resultados das pesquisas. As precauções adotadas na sua elaboração para impedir a violação do sigilo são ditadas meramente pelo bom senso comum, e não por estudos técnicos sérios. É muito provável que, em diversos desses arquivos, seja possível identificar os dados referentes a alguns indivíduos ou informantes. Um exemplo disso pode ser dado com a 1iberação de arquivos em fita contendo os·registros da. amostra de 25% do Censo Demográfico de 1980, Nesse arquivo, a identificação da localidade chega a município. Ora, como nos registros de pessoa se pode encontrar a ocupação dos indivíduos, em municípios onde a população for pequena não deve ser difícil identificar os registros referentes ao dentista, ao veterinário, ao prefeito da cidade caso eles tenham feito parte da amostra.

Assim, usando apenas uma pequena parte das informações contidas naqueles registros (por exemplo, o sexo, a idade e a ocupação) seria possível identificar o registro correspondente a algum indivíduo, e aí ter acesso aos demais dados ali contidos.

Fatos como esse só não ocorreram em abundância devido ao simples fato de que muito poucos usuários adquiriram cópia do arquivo da amostra de 25% dos registros do Censo Demográfico de 1980. Mas ele continua disponível para ser adquirido. Será que após o Censo

de 90 o quadro será mantido, mesmo na presença de·aumento da informatização·dos processos administrativos e decisórios possibilitado pela introdução maciça dos microcomputadores?

Outro aspecto que afeta a questão do sigilo é a metodologia da pesquisa: se as informações serão obtidas a partir de uma amostra ou da totalidade dos informantes, a definição da unidade informante e da unidade de investigação, e mesmo da metodologia de coleta - questionários auto-preenchidos ou preenchidos por entrevistador, devolução pelo correio, etc. Aqui, para tratar do assunto de forma adequada seria necessário detalhar a discussão, o que não parece oportuno neste documento.

Um último aspecto a mencionar diz respeito às variáveis pesquisadas, e sobre o sigilo·que se deve manter na divulgação. Há pesquisas que investigam algumas informações que são de domínio público: por exemplo, informações constantes do balanço das empresas, que são obrigatoriamente publicadas·em jornais de grande circulação. No entanto, ao proceder à desidentificação das tabelas para divulgação, todas as variáveis são omitidas para células com 1 ou 2 informantes à exceção do número de informantes. Assim, o IBGE mantém em sigilo informações que os próprios informantes são obrigados a divulgar·publicamente.

Em alguns casos, há associações que congregam firmas ou empresas, que divulgam anuários e outras publicações contendo dados individualizados a nível de empresa ou firma. Ao recolher esses mesmos dados, o IBGE se impõe a obrigação de mantê-los em sigilo.

Não se adotou até hoje a idéia de indagar aos informantes se desejam manter em sigilo os dados fornecidos ao IBGE. Idéias como essa poderiam contribuir para minimizar os problemas enfrentados na disseminação de informações. Há também que hierarquizar as variáveis para que o tratamento na disseminação seja corrente.

Com base nos exemplos citados, e na abordagem dos diversos problemas técnicos relacionados com a questão do sigilo das informações, pode-se perceber o grande atraso técnico do IBGE no tratamento dessa questão. Urge adotar providências capazes de eliminar esse atraso, de modo que o IBGE incorpore e passe a adotar procedimentos eficazes, de baixo custo, que propiciem aos técnicos, usuários e informantes segurança de que o sigilo das informações será mantido, onde couber, e não será violado apenas por simples descuido da instituição.

Na seção seguinte, serão introduzidos alguns conceitos técnicos relevantes para a compreensão do problema, e sugeridas algumas possíveis alternativas para investigação futura. Não se pretende esgotar o assunto, mas apenas indicar o caminho por onde se pode iniciar um trabalho de pesquisa mais profundo.

6. CONCEITOS E ALTERNATIVAS PARA PESQUISA

Os conceitos que serão introduzidos dizem respeito à divulgação de microdados. No entanto, podem ser aplicados para a divulgação de dados tabulados na situação em que ocorrem células com poucos informantes.

O primeiro conceito a ser introduzido é o de IDENTIFICAÇÃO. A identificação ocorre se uma relação de um para um puder ser estabelecida entre um indivíduo (ou

informante) conhecido em uma população ou subpopulação e um registro do arquivo que contém os microdados disseminados (ou uma célula da tabela que contém os dados divulgados). Uma conseqüência indesejável da identificação é a REVELAÇÃO: informações particulares do indivíduo (ou informante) identificado são reveladas.

A revelação é um problema porque consiste em uma violação do sigilo das informações, e está intimamente relacionada com a possibilidade de identificação de indivíduos ou informantes no conjunto de informações estatísticas disseminadas, incluindo as publicações usuais, fitas magnéticas, sistemas de acesso direto à Base de Dados, etc. A revelação se torna grave quando as informações descobertas são consideradas "sensíveis" pelos indivíduos ou informantes.

Como se pode perceber das definições dos conceitos, a identificação é um pré-requisito para a revelação.

De agora em diante, tratar-se-á da questão do sigilo das informações divulgadas por meio de arquivos em meios magnéticos contendo os microdados, isto é, dados a nível de cada informante referentes a diversas variáveis objeto da pesquisa que lhes deu origem. Os resultados assim obtidos poderão ser adaptados com relativa facilidade para o caso dos dados tabulados, onde o problema é menos grave, em princípio, devido à divulgação de menos informações que possibilitem a identificação dos indivíduos.

Levando em conta os conceitos de identificação e revelação anteriormente definidos, propõe-se repartir as informações contidas no registro correspondente a um informante em duas partes disjuntas: os dados de identificação, por um lado, e os dados confidenciais, por outro lado.

Os dados de identificação são constituídos pelas variáveis no registro que permitem a alguém identificar um registro, isto é, estabelecer uma correspondência biunívoca entre um registro do arquivo e um determinado indivíduo ou informante. Essas variáveis são denominadas variáveis chave ou de identificação. Algumas variáveis de identificação são bastante conhecidas, como o nome e endereço, mas há outras que podem ajudar·a identificar indivíduos como: idade, sexo, composição do domicílio, ocupação, raça, local de trabalho ou de moradia, etc. Nas pesquisas econômicas, a natureza da atividade econômica é uma variável de identificação.

Será admitida a hipótese de que as variáveis de identificação são todas categóricas, isto é, podem assumir apenas um número finito de valores.

Para se determinar se uma variável deve ser considerado como variável de identificação, deve-se levar em conta o conhecimento a priori ou a possibilidade de·conhecimento por outros indivíduos que não o próprio informante do valor dessa variável correspondente àquele informante, seja por meios próprios (vizinhança, parentesco, amizade, etc...) ou por meio de outras fontes de disseminação estatística. A presença desse conhecimento externo ou a priori por outros é fundamental para o tratamento do problema da identificação e da revelação.

A primeira implicação do reconhecimento da existência desse conhecimento prévio sobre as variáveis de identificação é que não se deve considerar essas variáveis como dados

confidenciais, no sentido de "informações que não devem ser reveladas pelo processo de disseminação das informações estatísticas".

Deste modo, é possível admitir que os conjuntos de informações de identificação e de informações confidenciais são disjuntos, isto é, mutuamente exclusivos. É possível, entretanto, que na prática surjam situações nas quais as informações disponíveis não possam ser separadas exatamente dessa maneira.

Admitir-se-á, daqui por diante, que a disseminação de informações estatísticas envolve, necessariamente, os valores de variáveis que pertencem ao conjunto de dados confidenciais dos informantes. De fato, se não fosse assim, a identificação jamais revelaria informações confidenciais e, portanto, a revelação nunca seria possível.

Em face dos conceitos já definidos, é possível estabelecer uma primeira regra básica para impedir ou dificultar a revelação. A regra é a seguinte: "Um arquivo de disseminação contendo microdados deve ser elaborado de tal forma que seja impossível para outros estabelecer corretamente ligações entre os registros individuais usando as informações de identificação do arquivo e o conhecimento prévio ou externo".

Percebe-se imediatamente que um elemento crucial do problema é o conhecimento prévio dos indivíduos sobre os outros: se ninguém pudesse conseguir qualquer informação sobre outro indivíduo, a identificação e, por conseguinte, a revelação seriam impossíveis. Assim, o risco da revelação depende da natureza e da quantidade de conhecimento prévio disponível ao interessado em fazer a revelação.

Um outro conceito de vital importância para entender como prevenir contra a identificação e a revelação é o conceito de UNICIDADE, que será introduzido adiante.

Primeiramente, denomina-se CHAVE uma variável ou conjunto de variáveis que possam ser usadas para a identificação dos informantes. Admitindo que a CHAVE pode assumir C diferentes valores, então as possíveis combinações de valores das variáveis componentes da chave podem ser denotadas por 1, 2 , ..., C.

O número de elementos da população com uma particular combinação de valores das variáveis da chave, por exemplo a combinação i , será denotado por F_i (i = l, 2 ,.., C). A soma $\sum_{i=1}^{C} Fi = N$ deve obrigatoriamente reproduzir o número total de unidades da população. Se a contagem do número de elementos com chave i for feita numa amostra dos elementos da população, o resultado deve ser denotado·por f_i (i = 1, 2, ..., C)

Definida esta notação, a probabilidade de que um e1emento selecionado ao acaso da população tenha o valor da chave igual a i é dada por $P_i = F_i/N$ i = 1, 2,..., C.

Daí, define-se a RESOLUÇÃO da chave por:

$$R = \left(\sum_{i=1}^{C} C\, P_i^2 \right)^{-1} = \frac{1}{\sum_{i=1}^{C} C\, P_i^2}$$

A resolução da chave é igual ao inverso da probabilidade de que dois elementos selecionados ao acaso da população tenham o mesmo valor da chave. Esta medida, a resolução da chave, fornece uma indicação do risco de identificação dos elementos da população com base nessa chave. Chaves cuja resolução é alta permitirão que muitos elementos fiquem isolados no arquivo de microdados, enquanto chaves com resolução baixa não permitirão que muitos sejam isolados dos demais.

Se todas as combinações de valores da chave fossem equiprováveis, isto é, se $P_i = 1/C$ $\forall i$, $i = 1, 2, ..., C$ então a resolução da chave seria $R = C$.

Deste modo, uma outra interpretação para o conceito de resolução da chave é o de "número efetivo de combinações" da chave.

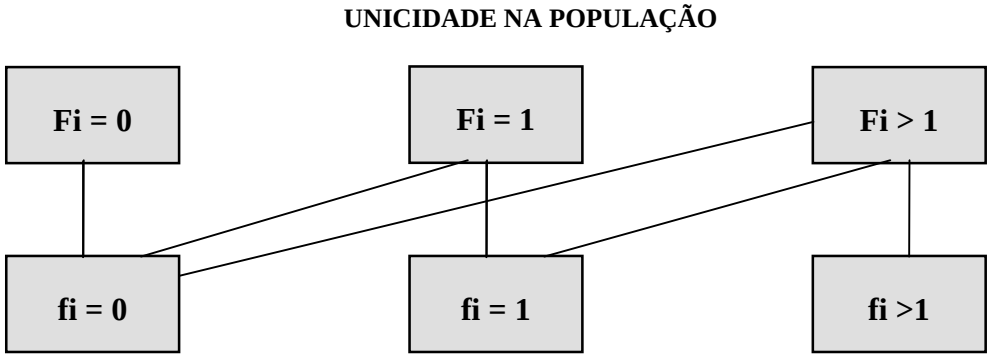
Agora já é possível introduzir o conceito de UNIDADE. Um elemento é ÚNICO na população se ele é o único a possuir uma particular combinação de valores da chave, isto é, se ele tem valor da chave igual a i e se $F_i = 1$, para algum i, $1 \leq i \leq C$.

De forma semelhante, um elemento é ÚNICO na amostra se ele possui valor da chave igual a i , com $f_i = 1$, para algum $1 \leq i \leq C$.

Qualquer elemento único na população será único na amostra, caso selecionado. Por outro lado, unicidade na amostra não implica unicidade na população. Deve ser notado também que unicidade na amostra pode ocorrer se apenas um elemento for selecionado entre os diversos elementos da população possuidores de um certo valor da chave.

Figura 1

Relação entre unicidade na população e na amostra



Agora fica evidente a influência do método de pesquisa na questão do sigilo: uma pesquisa por amostra pode revelar elementos únicos da população, se estes forem selecionados, e pode também conter elementos únicos apenas na amostra ($f_i = 1$ e $F_i > 1$) que podem ser identificados caso outros elementos consigam saber se os primeiros estavam selecionados ou não.

Para avaliar de forma precisa o risco da revelação, é necessário considerar diversos fatores intervenientes no problema:

a) conhecimento disponível sobre unicidade na população;

b) resolução de possíveis chaves;

c) quantidade de informações externas disponíveis para os possíveis interessados na revelação;

d) fração amostral e mecanismo de seleção da amostra, no caso de pesquisas por amostragem;

e) tempo decorrido entre o trabalho de campo, o período de referência das informações e a data da disseminação dos resultados (idade da informação).

Para levar em conta todos os aspectos mencionados, seria necessário imaginar uma particular aplicação, o que não parece oportuno fazer neste documento. Todo o capítulo 6 foi baseado no artigo de Keller, W.J. e Bethlehem, J.·G. (1987), onde os autores chegam a sugerir modelos e métodos que podem ser adotados em certas situações particulares para dar conta do problema de avaliar o risco de revelação.

A intenção predominante ao escrever este capítulo foi trazer à tona o lado técnico/metodológico do problema, ilustrar sua complexidade, a fim de que as recomendações

no sentido de realização de um esforço de pesquisa sobre esse assunto sejam melhor entendidas e avaliadas.

O problema técnico da revelação já foi tratado em inúmeros artigos, que se precisa conhecer melhor e analisar com profundidade, para que se possa alcançar os fins desejados: identificar, adquirir e disseminar uma metodologia segura, barata e eficaz para tratar o problema de preservação do sigilo das informações estatísticas.

Apenas para ilustrar o problema da unicidade, retirou-se do artigo de Keller e Bethlehem (1987) um exemplo impressionante: dos 83.799 domicílios numa certa região, 23.485 são habitadas por famílias compostas de pai, mãe e dois filhos. Se as variáveis da chave de identificação fossem idade do pai, idade da mãe, e idade e sexo dos dois filhos, 16.008 dos 23.485 domicílios dessa região são únicos com relação à chave escolhida, o que dá uma proporção de 68% do total de domicílios com famílias como as que foram escolhidas!

É uma pena que avaliações como essa não façam parte da rotina de preparação dos produtos de disseminação de uma pesquisa qualquer do IBGE, ou não?

O trabalho de pesquisa deveria abranger a parte metodológica, no sentido de identificar e adquirir os modelos necessários para estimar a resolução de uma chave num arquivo de dados qualquer, e também uma parte empírica, que consistiria num levantamento abrangente da situação do sigilo nas diversas pesquisas e inquéritos do IBGE.

7. CONCLUSÕES E RECOMENDAÇÕES

De tudo que foi exposto nas seções anteriores, emergem como principais conclusões:

a) o tratamento da questão do sigilo é precário e heterogêneo nas diversas áreas do IBGE, não só com respeito aos meio e técnicas empregados como também com relação aos resultados que obtém; 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 á rotina das pesquisas; alguma 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,1egais e éticas que devem ser orientadas 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 vista das conclusões acima, e considerando a oportunidade oferecida pela realização do Seminário de Disseminação de Informações, apresentam-se para consideração as seguintes recomendações:

(i) aproveitar o Seminário de Disseminação de Informações para ampliar e aprofundar o debate em torno da questão do sigilo, principalmente nos seus aspectos políticos, legais e éticos;

(ii) criar, a partir do Seminário, um grupo de trabalho composto por técnicos das áreas de produção e disseminação, encarregado de realizar a pesquisa dos métodos e técnicas existentes para o tratamento da questão do sigilo, adaptá-los para uso no IBGE, disseminá-los nas·áreas envolvidas com o problema e propor a normatização dos procedimentos a adotar com vistas à manutenção do sigilo das informações; deve ser ressaltado que um grupo de trabalho com essas atribuições e características deveria ocupar seus integrantes em tempo integral por·um período de 6·a 8 meses para apresentar resultados satisfatórios;

(iii) envolver, a partir do grupo de trabalho criado, todos os departamentos e áreas da organização relacionados com a questão, mediante treinamento de seus funcionários,

do realce da questão do sigilo no planejamento metodológico das pesquisas, e mediante levantamento metódico da situação vigente em cada uma das pesquisas;

(iv) promover, caso necessário, os esforços exigidos no sentido de atualizar a legislação vigente sobre o assunto;

(v) encarar as soluções propostas como capazes de produzir resultados mais a médio e longo prazo do que de imediato; as principais vantagens serão medidas através do nível de credibilidade da organização, da cooperatividade dos informantes, e principalmente da satisfação dos usuários com os resultados divulgados.

8. BIBLIOGRAFIA

01. KELLER, W.J. e BETHLEHEM, J.G. **On Disclosure Control of Micro Data**. Netherlands Central Bureau of Statistics. Department for Statistical Methods. BPA no. 7324-87-M1; 1987.

02. SILVA, P:L.N. **O Problema da Desidentificação dos Dados das Pesquisas da SUICOM - Propostas para Solução**. IBGE - Setembro de 1986 - - Documento não divulgado.

03. CONFE - Conselho Federal de Estatística - Legislação Básica - Estatístico e Técnico em Estatística de Nível Médio. 1977.

Destaco que uma primeira relação de bibliografias a serem consultadas para iniciar o trabalho de pesquisa sobre o tema já foi elaborado, e pode ser encontrada na bibliografia número 2.