

1. Introdução;
2. A necessidade da associação de dados e da visão do detalhe;
3. Sistema e usuário: razões para uma falta de entendimento;
4. Acesso e análise de dados: o papel do usuário;
5. A retaguarda histórica;
6. A retaguarda cultural;
7. Os elementos estruturais;
8. Matchings;
9. Sigilo e privacidade;
10. Conclusão.

Graciano Sá\*\*

## CADASTROS, TABELAS E PADRÕES: UM EXAME DO SISTEMA ESTATÍSTICO NACIONAL\*

### 1. INTRODUÇÃO

Este artigo faz uma análise do Sistema Estatístico Nacional no sentido de propor que as atuais práticas metodológicas e industriais do sistema não satisfazem às exigências do usuário moderno e precisam de uma mudança.

O artigo enfoca a necessidade pelo usuário do detalhe da informação e, pois, do acesso a dados geralmente dispersos entre edições de arquivos e de agências do sistema estatístico; discute a forma do relacionamento existente entre sistema e usuário e analisa o objeto industrial que concretiza aquele relacionamento; aborda a equivocada percepção que o sistema tem do que é acesso a dados e análise de dados e mostra como estes conceitos distinguem usuários diretos e indiretos do sistema; analisa os aspectos históricos, culturais e tecnológicos subjacentes; conclui pela necessidade do reexame do objeto industrial do sistema e pela integração dos arquivos de dados de seus diferentes agentes.

A pretendida revisão de procedimentos aponta para novos rigores conceituais na preparação de arquivos de dados, para o uso difundido do exercício de *matching* entre arquivos e para a necessidade da montagem dos elementos estruturais fundamentais, que estão faltando ao Sistema Estatístico Nacional, os cadastros, tabelas e padrões.

Observações finais levantam a questão dos efeitos desta revisão para a liberdade do indivíduo, seu sigilo e sua privacidade.

25

### 2. A NECESSIDADE DA ASSOCIAÇÃO DE DADOS E DA VISÃO DO DETALHE

Pondere o leitor preliminarmente sobre questões como as seguintes, a título de exemplo:

1. O serviço da receita federal deseja estudar uma alteração no imposto de renda das pessoas físicas, com vistas a determinar o reflexo sobre a renda disponível das famílias e sobre a receita global do tributo.
2. O IBGE quer determinar se a apuração de informações sócio-econômicas rurais pode ser feita por via de uma amostra de estabelecimentos agropecuários, de forma a reduzir o custo e o tempo de processamento de censos agrícolas.
3. A Eletrobrás preocupa-se com a perda do valor real da tarifa de energia elétrica e deseja estudar a elasticidade da demanda dos bens do orçamento industrial e domiciliar associada ao preço da energia.
4. A Secretaria de Planejamento quer conhecer a relação preço/custo de estabelecimentos industriais em períodos de expansão de demanda, de forma a identificar os setores que comprimem a margem e os que aumentam preços mais que custos, propagando a inflação.
5. O Ministério de Minas e Energia quer avaliar o balanço energético de biomassas, para certificar-se de

\* O artigo beneficiou-se da crítica dos Profs. Isaac Kerstenetzky e Mario Dias Ripper a quem o autor agradece. As opiniões expressas no texto são da responsabilidade exclusiva do autor e não refletem de forma direta ou indireta posições da instituição com a qual o autor está associado.

\*\*O autor é engenheiro pela Universidade Federal do Rio Grande do Sul e mestre e doutor em *management* pelo Instituto Tecnológico de Massachusetts, EUA. Presentemente é diretor adjunto da Diretoria Técnica da Fundação IBGE.

que o transporte daqueles materiais, diante dos meios e acessos disponíveis, não consome a contribuição positiva da conversão em energia, no ponto de transformação.

6. Uma organização bancária nacional precisa conhecer a distribuição espacial e a composição sócio-econômica de domicílios em certos centros urbanos brasileiros, para determinar a localização correta de novas agências.

7. A Prefeitura de uma cidade brasileira, com evidência de acentuado índice de urbanização, precisa defender junto ao Governo federal o financiamento da ampliação do seu sistema de transporte, de saúde e de saneamento, tendo por isso de analisar o efeito dos acréscimos de população naqueles serviços públicos.

Questões como estas têm em comum, invariavelmente, os atributos seguintes: *a)* são reais; *b)* pertencem a uma coleção quotidiana de indagações semelhantes, atravessando diversos planos de gestão, em nível operacional, tático e estratégico, e cuja formulação torna-se cada vez mais complexa na elucidação do comportamento da pobreza, educação, saúde, desemprego, urbanização, mercados, etc.; *c)* a solução requer dados, dentro do Sistema Estatístico Nacional, que não pertencem exclusivamente ao mesmo agente estatístico, ou nele ao mesmo inquérito estatístico, ou nele à mesma edição do mesmo inquérito; *d)* o objeto de atenção requer exame do menor detalhe de investigação, portanto do detalhe de investigação do documento alimentador do dado, a nível de pessoa, família, estabelecimento ou item; *e)* são questões de análise estatística, não de inteligência (no sentido investigatório, de vigilância, atribuído ao termo). Têm por isso um objeto de exame que não precisa de identificação explícita, bastando que seja reconhecido como uma entidade válida de observação, dentro do gênero de fenômeno em análise; *f)* são questões que deveriam ser resolvidas dentro dos recursos do Sistema Estatístico Nacional.<sup>1</sup>

O que principalmente têm em comum estas questões é o fato de serem insolúveis, sem o recurso à associação de dados entre agências do sistema estatístico, ou entre arquivos de inquéritos de diferentes funções ou edições na mesma agência daquele sistema. Uma associação propositadamente construída entre registros de diferentes arquivos de dados designa-se *matching*, ou casamento.

A primeira das questões que formulamos, quanto ao efeito de alterações tributárias na renda disponível das pessoas, requer a construção de uma base de dados onde registros de diferentes fontes, por exemplo, de arquivos fiscais (estrutura da receita e da despesa da família), censitários (estrutura, idade, escolarização da família) e previdenciários (mobilidade e emprego na família), são casados, família a família, para análise das conseqüências de diferentes hipóteses de tributação. A segunda questão, de amostragem em censos agrícolas, resolve-se pelo exame de arquivos de censos passados com o fito de colher-se a amostra que melhor simula os predicados funcionais

(tipo de exploração econômica) e regionais (estrutura, dimensão, capital e trabalho) dos estabelecimentos do universo. A terceira questão impõe a construção de um modelo de que participem dados da estrutura do fornecimento de energia elétrica, a partir dos cadastros das concessionárias, casados a dados censitários da estrutura do consumo familiar em nível regional, e de insumos e produtos de empresas industriais e agrícolas, bem como seus preços.

Para a solução das demais questões, não será difícil observar que também os elementos necessários à análise constam de diferentes edições de um mesmo arquivo de dados ou de edições equivalentes de arquivos de agências distintas do Sistema Estatístico Nacional. Por isso, o Sistema Estatístico Nacional é aquele conjunto de mecanismos, rotinas e dados que tem por finalidade fornecer os elementos para a solução daquelas entre muitas outras questões. Se for inexecutável realizar-se o *matching* de arquivos, por carência do recurso a cadastros, tabelas e padrões do Sistema Estatístico Nacional, as questões tornam-se novos exemplos de problemas cuja solução irá exigir grande esforço de imaginação ou, em casos relevantes, levantamentos de dados especiais, *ad hoc*, à margem do sistema. Esta tem sido normalmente a direção para a qual envereda o usuário do Sistema Estatístico Nacional, porquanto é reconhecido e tem-se como lugar comum que a parte mais difícil na solução de qualquer problema analiticamente formulado é a angariação dos seus dados. Dados quase nunca estão disponíveis a tempo e hora ou, quando estão, não são de acesso fácil na forma e nível de detalhe desejados, ou não são estritamente comparáveis entre si, ou padecem de outra deficiência do gênero.

Dados em nível de detalhe crucialmente menor, contudo, tornam-se os instrumentos que o usuário tem para contrapor-se à complexidade dos fenômenos do momento atual. Extrapolações do comportamento de agregados por via de proporções, elasticidades e uma ou outra relação multivariada já não satisfazem. O concurso da análise é agora multidisciplinar e o exercício da elucidação de questões sócio-econômicas desenrola-se num plano onde o efeito de cada variável tem de ter seu peso devidamente medido. A regularidade previsível de agregados tem sido tradicionalmente útil para substanciar decisões de âmbito estratégico, por exemplo, na definição de uma política demográfica para o país ou na determinação da forma preponderante de geração de energia termoelétrica no próximo século, se nuclear ou a carvão mineral. Seria impróprio para o Governo, entretanto, e poderia tornar-se fatal para a empresa, respaldar apenas em agregados estatísticos, na época em que vivemos, as decisões que pertencem a níveis táticos e operacionais de ação, como as de contenção do êxodo rural ou de determinação do local, capacidade e início de operação de futuras usinas termonucleares. A dificuldade que temos em dar curso detalhado à solução de questões deste tipo, com apoio no Sistema Estatístico Nacional, denota que este sistema é, sob inúmeros aspectos, deficiente. Qualquer sistema estatístico, até nos mais avançados países, será deficiente, no sen-

tido de não poder antecipar plenamente as exigências de informação do Governo e das empresas para o controle das suas operações. Não obstante, certas propriedades de um sistema estatístico, encontradas em outros países, poderiam sem muito esforço ser incorporadas ao nosso.

Neste particular, uma leitura nos relatórios Kaysen<sup>2</sup>, Ruggles e Dunn<sup>3</sup> ao Congresso dos EUA em 1966, por ocasião do debate despertado naquele país quando da proposição de um Centro Nacional de Dados, há 15 anos portanto, deixa o leitor com a sensação de estar revendo um diagnóstico da atual situação estatística brasileira. Pois as deficiências que hoje se observa no Sistema Estatístico Nacional foram, na época, e têm sido, desde então, repetidamente expostas. O entendimento daquelas advertências não foi, entretanto, inteiramente absorvido em nosso meio.

O que há de errado com o Sistema Estatístico Nacional? Basicamente o fato de não dispor, ou não expor, a tempo e hora, os dados de que se precisa<sup>4</sup>.

A dificuldade de adequar o sistema estatístico de nosso País aos padrões exigidos para o planejamento de uma sociedade em crescimento, nos limites da escassez, tem raízes históricas, tecnológicas e culturais. Proponho, na seqüência, discutirmos as dificuldades que inibem a realização do Sistema Estatístico Nacional e o que nos cabe fazer para restaurar-lhe a função.

### 3. SISTEMA E USUÁRIO: RAZÕES PARA UMA FALTA DE ENTENDIMENTO

Na trilha de sistemas estatísticos de outros países, o nosso tem problemas de adequação, no sentido do grau em que atinge seus objetivos; de eficiência, em termos do volume do *output* que produz por unidade de *input*; e de eficácia, no que diz respeito à qualidade do seu *output* por unidade de *input*.

Os problemas de adequação referem-se principalmente ao *tempo* e à *forma* de acesso a dados. Tempo (velocidade) e forma de acesso são atributos importantes da informação e, juntamente com outros — como unicidade, coerência e associatividade — dão à informação a medida de sua utilidade. Tempo e forma, entretanto, têm sido levados, em nosso País, a reforçar desnecessária e mutuamente a inaptidão do sistema estatístico a seus objetivos. Com efeito, o usuário atual deseja que aquele sistema lhe ofereça acesso rápido a dados descomprometidos de agregação e análise prévia; portanto, sob forma de registros magnéticos no nível primário da coleta do inquérito (microdado) e com qualidade numérica certificada por padrões conceituais estáveis, intra e interagências. Entretanto, ele obtém: *a*) acesso lento: dois a cinco anos, em média, para os inquéritos mais importantes e volumosos do sistema; *b*) acesso em doses homeopáticas, normalmente regionais; *c*) acesso a dados agregados em forma tabular; *d*) acesso a meio impresso, não legível por computador. Tempo e for-

ma de acesso a dados, nestas circunstâncias, roubam ao usuário muitos graus da liberdade de que ele necessitaria e poderia ter em sua capacidade de produzir informações. Isto, entretanto, não é tudo. O usuário ganha, ainda, dados normalmente descomprometidos de coerência, por ser comum em nosso meio a necessidade que tem de conferir com grande cuidado notas e páginas introdutórias, em busca das variações conceituais que tornaram *este* inquérito diferente, em grau variável, dos de mesma família, edições anteriores, ou de outras famílias, na mesma edição.

A apresentação de estatísticas agregadas em forma tabular tem sido repetidamente condenada como objeto de sistemas estatísticos<sup>5</sup>. Ela só facilita a recuperação (*retrieval*) de formas preconcebidas de informação, cerceando a análise em face da inviabilidade da associação de dados primários. Tem sido também condenada por revestir-se de um anacronismo derivado de épocas em que a tecnologia era consideravelmente mais pobre, subordinando o usuário à síntese que o sistema lhe fornecesse. E por inculcar o erro de então de que o dado primário não tem valor fora do sistema, consentindo-se na sua mutilação ou destruição.

Os problemas de eficiência, por sua vez, dizem respeito à desproporção entre o que entra no sistema estatístico (dados, sobretudo, mas também recursos e dinheiro) e o que dele sai. De um lado, a multiplicidade de agentes processadores convida a redundâncias de *input*, onerando o sistema e o respondente; de outro, mesmo no limite da tecnologia, são necessários meses de trabalho para digerir *inputs* de bilhões de caracteres/ano, ordem de grandeza da carga do nosso sistema estatístico. O sistema, então, congestionase na difícil tarefa de limpar e certificar arquivos, algo que se agrava quando o seu objeto industrial não é reconhecido como o arquivo magnético primário, limpo e certificado (a base de dados), mas a síntese impressa que dele resulta. Um sistema que, premido pelas novas exigências da sociedade, em absorver volumes crescentes de entradas, quando cumpre o ritual de liberar a publicação programada deixa atrás de si um rastro de concessões técnicas sem tempo para a estabilização de métodos e padrões. Em épocas particularmente difíceis, perde a confiança, e não publica.

Os problemas de eficácia, finalmente, são os que decorrem da pobreza do *matching* intra e interagências, em nível de microdado.

Microdado é a imagem magnética do instrumento da coleta de dados, o único a partir do qual todas as sínteses podem ser construídas. Os problemas de eficácia se pronunciam porque o sistema não tem compromissos com o passado e por isso não vincula, por cadastros, tabelas e padrões, as associações das bases de dados do futuro. Num sistema sem história, cada inquérito entra em fabricação como objeto novo, guardando semelhanças (*comparabilidades*) com objetos da mesma família, sem, entretanto, possuir as estruturas de dados alinhadas como preceito de método, perdendo compatibilidades essenciais em *mat-*

*chings* campo a campo, em nível de microdado. Fica então comprometida, em certa medida, a qualidade do dado que o sistema finalmente produz. Estas propriedades do Sistema Estatístico Nacional (quanto à eficiência, à eficácia e à adequação) interpõem-se no caminho do entendimento entre sistema e usuário.

O primeiro não alcança o segundo, talvez por não reconhecer onde ele verdadeiramente se encontra. A percepção que o sistema tem sobre quem é o usuário é prejudicada sobretudo por um equívoco conceitual entre o que é acesso e análise de dados, como discutiremos a seguir.

#### 4. ACESSO E ANÁLISE DE DADOS: O PAPEL DO USUÁRIO

Os problemas anteriormente mencionados do Sistema Estatístico Nacional originam-se, fundamentalmente, do entendimento que o sistema tem de que o seu objeto industrial é a síntese tabular impressa e, o usuário, um dependente. Com efeito, enquanto o que resultar do sistema for apenas um programa de tabulações, os rigores da comparação campo a campo, em quesitos inquiridos, poderão ser afrouxados, de vez que o sistema tem por ofício garantir a comparação entre acumuladores de arquivos (os agregados) e não entre campos de registros de arquivos (os itens). E, enquanto o usuário for um dependente, toda a solicitação especial que vier a exigir, em nível de microdados, poderá ser atendida pelo sistema, ainda que com restrições, pelo emprego da inteligência (*software*) que só este possui, um processo de reconstrução que o usuário não poderia fazer de *per si*, e que reforça o seu grau de dependência no sistema. Nesta relação de dependência, o sistema não expõe o dado, quando o possui à espera de que o usuário o solicite. O sistema então não fornece ao usuário acesso (*retrieval*) ao dado explicitadamente desejado, conforme sua solicitação, mas análises de dados, com o que o usuário tem de contentar-se, normalmente depois de alguma espera.

A diferença entre acesso a dados e análise de dados não é facilmente compreendida e reside no centro do papel que o usuário do sistema estatístico tem para desempenhar. De acesso precisa quem realiza uma análise; de análise, quem decide sobre uma resposta.

Escalões superiores, no Governo e na empresa, precisam de análise (embora normalmente pensem em termos de acesso), pois é neste domínio que se situam respostas às questões como as que abriram este texto. Escalões dirigentes são, por isso, usuários indiretos do Sistema Estatístico Nacional.

O usuário direto daquele sistema é normalmente um intermediário entre um nível de decisão (um nível dirigente) e uma base de dados (o sistema estatístico), a quem se cometeu a tarefa de extrair informação da massa de dados disponível. A ele só interessa acesso a dados que mantenham abertas as opções de análise.

Para qualquer destes participantes, dirigentes no Governo e na empresa e analistas usuários do Siste-

ma Estatístico Nacional, análises e sínteses previamente incorporadas a dados, apresentados em forma tabular, são contribuições despojadas das peculiaridades do problema em questão, prejudicando o seu livre equacionamento. É nestes termos que o usuário naturalmente se ressentia da dependência que o sistema ainda lhe impõe.

A restauração do Sistema Estatístico Nacional reside, portanto, de início, na remoção daquela relação de dependência, ou seja, no duplo reconhecimento, primeiro, de que o seu mais fértil e rigoroso usuário é quem transita entre a base de dados e o nível dirigente, num leva-e-traz entre acesso e análise de dados; e, segundo, que para ampliação do efeito multiplicador daquele usuário, o objeto da indústria do sistema tem de ser as bases de dados que o sistema constrói, através da certificação e consistência de arquivos magnéticos primários, imagem dos questionários da coleta e de documentos alimentadores em geral.

Esta constatação impõe novas e profundas restrições no funcionamento industrial das agências do sistema estatístico, porque revelam-se substancialmente diferentes os métodos de trabalho voltados para sínteses impressas e os que asseguram alinhamento e comparabilidade de dados e nível de registros magnéticos primários.

Tal reconhecimento situa também em plano muito distinto o relacionamento do sistema com seu usuário, porque o sistema fica liberado do compromisso de produzir sínteses impressas, exceto onde estritamente cabível, podendo voltar-se para o atendimento de suas múltiplas responsabilidades previstas em lei, e o usuário é induzido a realizar a análise com independência.<sup>6</sup>

O encaminhamento do Sistema Estatístico Nacional nesta direção depara-se, entretanto, com obstáculos que, além de tecnológicos, são predominantemente humanos. As áreas de resistência encontram força em posições históricas e culturais. Estas são questões para as quais voltamos a seguir.

#### 5. A RETAGUARDA HISTÓRICA

A separação exigida, nos planos em que se encontram o Sistema Estatístico Nacional e seu usuário atualmente, é outro fruto das tensões de uma época em que a rapidez das mudanças tecnológicas e sociais avassala o indivíduo e o impede de acompanhá-las. Foi-se o tempo em que o ritmo dos eventos ajustava-se ao consumo do capital intelectual de uma pessoa adquirido nos anos formativos e na experiência que colhesse em vida. Também as instituições, cuja adaptação às mudanças vincula-se à capacidade de ajustamento dos indivíduos, não acompanham mais com rapidez as acomodações impostas por fenômenos em curso.

Observe-se, a respeito, os esforços do Sistema Estatístico Nacional para alinhar-se com o ritmo do progresso nestes últimos 20 anos. A fase moderna do sistema, pode-se dizer, começou em 1960, quando o

IBGE adquiriu um computador de grande porte, entre as máquinas de maior capacidade na época, muito desproporcionado entretanto às condições técnicas então vigentes no País. O Univac 1.105, nome daquele computador, um colosso de válvulas, incompreendido e mal dimensionado, acabou por mergulhar sua agência em prolongado declínio técnico do qual ela só conseguiu emergir 10 anos depois. Entre 1960 e 1970 e boa parte dos anos mais recentes, o Sistema Estatístico Nacional esforçou-se por dominar uma nova tecnologia, ocasião em que estruturou os principais birôs estatísticos e administrativos do Governo: Serpro (1964), Fundação IBGE (1967), Prodesp (1969), Prodasen (1972), Dataprev (1974), entre os mais conhecidos. Contudo, ainda que por lei estas sejam agências de um mesmo sistema estatístico, a função de cada uma é divisional e específica, normalmente distante dos propósitos de uma ação estatística comum. A qualidade dos métodos que as agências isoladamente empregam mostra-se também heterogênea, conseqüência de ser heterogênea a retaguarda técnico-cultural dos quadros humanos com que contam. Por isso, o sistema estatístico apóia-se numa rede de agentes processadores dirigidos, de momento, para o domínio da tecnologia em nível divisional, e às voltas com a uniformização da qualidade técnica do seu pessoal. Não se pode esperar, em conseqüência, que contribuam para a formação de um sistema estatístico mais aceleradamente do que vêm fazendo.

As agências do sistema estatístico atingiram nestas duas décadas, com mais velocidade nos últimos anos, alguns processos na estruturação de dados primários. Com respeito à formação de cadastros e âncoras de *matching*, são exemplos os cadastros CGC/CPF, PIS/Pasep/Rais, Código Brasileiro de Ocupações, Código de Municípios, etc.; com respeito a padrões de arquivamento, deve-se a uma ação do Ministério de Minas e Energia muito da situação de ordem em que se encontram os dados de produção e consumo de energia elétrica no País;<sup>7</sup> e no fornecimento de arquivos magnéticos para pesquisa, é conhecida a fita 1/100 do censo demográfico de 1970, e a disponibilidade de dados selecionados dos cadastros de contribuintes da receita federal. Estes exemplos, contudo, não têm origem no propósito de uma inteligência central (o sistema estatístico) e decorrem mais de ações espontâneas provocadas pela complexidade da máquina administrativa do País (CGC/CPF), da abrangência de novos diplomas sociais (PIS/Pasep/Rais) e estímulos semelhantes. São exemplos de que o usuário acompanhou a tecnologia e pressionou no sentido da modernização do sistema, mas também são exemplos de iniciativas estritamente divisionais acompanhando a impulsão da tecnologia.

As massas de dados, custodiadas por estes agentes, apenas remotamente são entendidas como integrantes de um sistema comum e por isso são construídas sem compromissos de *matchings* interagências. Âncoras de *matching*, criadas por certos agentes do sistema, não são uniformemente adotadas pelos demais. O perturbador é que não são aceitas uniformemente nem mesmo entre seções do mesmo agente,

resguardando cada parte independência em suas estruturas funcionais de dados.

Historicamente, em nosso meio, não há garantia de padrões estáveis em estruturas de dados e são inúmeros os exemplos onde a lei de formação de certos códigos fundamentais altera-se com freqüência. As agências, e nelas as seções, atuam então em subsistemas estanques, sem realizar o alcance que teriam suas ações se, por campos de dados comuns, pudessem ser casadas com ações de outros agentes e seções do sistema.

Mantido o isolamento divisional dos agentes do sistema estatístico, amplia-se o vácuo entre ele e o usuário. Por isso, os progressos em equipagem, conseguidos em 20 anos, para construir um sistema estatístico à altura dos que existem em outros países, não foram suficientes, ainda, para dar ao nosso a organicidade que lhe está faltando. Será difícil dar integração ao Sistema Estatístico Nacional pela articulação administrativa do trabalho de seus múltiplos agentes, porque o objetivo divisional de cada um será levado a prevalecer. A ação independente das partes de um sistema, contudo, é a negação da existência do sistema, por definição. Importante por isso romper-se esta tendência histórica e escrever-se o manual das regras do jogo do *matching* de dados entre as agências do sistema, para que cada uma possa cumprir a sua função, sem perdas irremediáveis para o conjunto.

29

## 6. A RETAGUARDA CULTURAL

Muito da distância que separa o sistema estatístico de seu usuário deve-se também ao choque entre as culturas técnicas e administrativas das agências do sistema estatístico, decorrente da distinção das atitudes derivadas do empirismo gerencial e político dos *managers*, de um lado, e da racionalidade objetiva dos analistas de sistemas do outro. Os estereótipos de uma e outra cultura têm sido profusamente discutidos na leitura. Por exemplo em Churchman:<sup>8</sup>

“Much of Management Science has been conducted under the very naive philosophy that a certain kind of reason must prevail and that once this has been made clear the manager will accept it or be charged with gross negligence or gross stupidity...”

Many scientists think they are wooing management, that management is the coy lady, that management must be ‘sold’. Such attitude is naive: management is aware of wath the scientist is trying to do and has made up its mind about how science should be used, namely to tackle the important but nonetheless uninteresting problems of technology in government and industry.”

Ou Argyris:<sup>9</sup>

“Historical emphasis upon power over competence and fuziness over explicitness has staffed corporations with managers with qualities and competence

different than those required to manage with information systems.

The new skill that may be required is for the manager to deal with the interdependence among facts..."

O analista é um racional acadêmico que almeja um mundo estável, integrado, institucional. A expressão do seu pensamento realiza-se, quase sempre, na palavra escrita, se possível, pela formulação explícita do modelo da situação. Em seu subconsciente encontram refúgio critérios definidos de otimização de algum valor tangível ou *utilidade*, uma função-objeto através da qual ele pode escolher alternativas e dar evidência da coerência de seu pensamento. O *manager* é, entretanto, um racional pragmático, cuja visão volta-se para um mundo variado, autônomo, pessoal. Deseja ter abertas as suas opções, habita um mundo intensamente verbalizado, procura tornar os objetivos da organização congruentes aos seus e por isso percorre caminhos cuja coerência só pode ser entendida pelos que conviveram pelo menos uma vez nos corredores do *management*. Seu subconsciente abriga critérios subjetivos de otimização. Diz-se que ele maximiza *satisfação*.<sup>10</sup>

Estas duas culturas estão em permanente conflito, tão distintos os seus pontos de referência quanto mais nítidos os clichês. Quem sofre as conseqüências do embate são os sistemas que contêm ambas as partes. Aqueles sistemas têm curso, integrados ou fragmentados, conforme a influência residir em uma ou outra área. Geralmente, a influência confunde-se com o poder da última palavra.<sup>11</sup>

É freqüente, então, ter-se a análise dos sistemas recuada a nichos técnicos passivos, onde abdica de interpelar por normas amplas e comuns de estruturação, arquivamento e processamento de dados. Onde o poder da última palavra do *manager* é o constante argumento de persuasão, e o analista contenta-se em codificar em computador o pensamento exclusivo daquela outra parte deseducada de suas conseqüências, os efeitos nas fragmentações de sistemas podem ser consideravelmente duradouros.

Nada mais aberto, tem-se dito, que a mente de uma criança ou a memória não programada de um computador, nem mais obstruído que os hábitos de um homem idoso ou de um computador programado. Para dar harmonia às partes componentes de um sistema organizacional, em busca da integração dos seus dados, é importante arbitrar-se o embate das culturas técnicas e administrativa dentro de cada agência da organização, porque é naquele nível que se sacrificam coerências importantes para a associação de bases de dados divisionais.

Neste antigo confronto cultural não existem, naturalmente, vencedores nem vencidos. Como um pêndulo, a autoridade na organização progride de um a outro campo, em intervalos que dependem das inércias administrativas construídas pela fragmentação ou pela integração dos sistemas.

Para uma análise das decisões que afetam o deslocamento da autoridade, de uma para outra área, o

importante é poder-se contar, em cada caso, com a experiência dos que conviveram com as verdades de ambos os lados. Não são muitos, entretanto, os que possuem a experiência que está faltando. Em nosso País, os meios técnico e administrativo contêm gerações de pessoas cujos paradigmas são profundamente distintos, por terem sido apanhadas no torvelinho das mudanças dos últimos anos, circunstância que aparece visivelmente em quase todo lugar e, certamente, dentro das agências do Sistema Estatístico Nacional.

Neste sistema, particularmente, o pragmatismo político dos *managers* muito contribuiu para a fragmentação divisional do trabalho das suas agências e, dentro de cada uma, para a fragmentação funcional dos subsistemas. É preciso repensar o acerto desta posição.

Tudo leva a crer que o progresso do País, aumentando o lastro técnico-cultural de suas camadas dirigentes intermediárias, tenderá agora a dirigir o pêndulo das decisões no caminho da integração de sistemas estatísticos. A integração de sistemas de dados em geral depende, entretanto, do rigor metodológico que for aplicado em elementos estruturais fundamentais daqueles sistemas. Este é o objeto de nosso próximo assunto.

## 7. OS ELEMENTOS ESTRUTURAIS

Quando Richard Ruggles e Karl Kaysen propuseram, em 1966, ao governo americano, um centro nacional de dados, a idéia contrapunha-se, na época, à ação descentralizada e dispersiva das agências de dados daquele país "dilapidando um recurso nacional". O objetivo do centro era a custódia física e lógica de milhares de fitas magnéticas, de forma a permitir o seu acesso desde um ponto central. Na análise que fez da proposição, foi Edgar Dunn quem situou o problema corretamente:<sup>12</sup>

"In a fundamental way file accessibility is the issue of file compatibility, which is inseparable from production practices that determine the organization and quality of the file. Steps that can be taken to improve file compatibility and accessibility will substantially increase the effective scope and utility of present files without a change in the size of the files or the attributes of the respondents they contain... The central problem of data use is one of associating numerical records... It rests upon the file being constructed on the basis of standards, maintained in effective condition, and serviced by institutional arrangements and a technical system capability, that will allow it to be reprocessed and combined successfully with other records in a wide variety of ways that will meet the special requirements of a wide range of users."

Em síntese, um sistema estatístico nacional pode constituir-se, ele mesmo, no análogo a um centro nacional de dados, se houver associatividade entre registros magnéticos dos arquivos divisionais.

A necessidade de institucionalizar um centro nacional de dados fica, então, muito reduzida se, desde um ponto central, normativo, estabelecerem-se os elementos da associação de registros de dados intra e interagências. Para este objetivo são necessários os elementos estruturais essenciais, que são cadastros, tabelas e padrões.

Cadastros são *universos dos objetos de pesquisa*, por exemplo, a lista das empresas industriais do País. Cadastros preocupam-se em reter os descritores mínimos dos objetos, principalmente *quem* é o objeto e *onde* ele se encontra. Bases de dados distinguem-se de cadastros por conterem os atributos medidos em cada objeto, em cada inquérito e, naturalmente, o identificador que associa o objeto, no cadastro, aos registros de suas medidas na base.<sup>13</sup>

Tabelas são *organizações de referência*, que asseguram coerência funcional ou espacial entre edições da mesma base ou entre bases diferentes da mesma edição. Tabelas contêm entidades ou códigos universais, por isso presentes longitudinais e transversalmente em bases de dados. Por exemplo, as tabelas funcionais de produtos industriais, de atividades econômicas e de ocupação e as tabelas espaciais de estado, município, setores censitários, endereçamento postal e logradouros.

Padrões são *regras de estruturação de dados* em nível de campo, em registro magnético, estabelecendo o seu alinhamento com outras ocorrências do mesmo campo entre inquéritos, edições de inquéritos e agências processadoras. Padrões são esmiuçados em dicionários de campos de dados, contendo a conceituação do dado, sua estrutura, a interpretação de caracteres e dígitos na estrutura, forma de gravação, e outros descritores técnicos relevantes.

Cadastros, tabelas e padrões conferem estrutura a bases de dados. Informação, tem-se dito, é o resultado da união de dados e estrutura, e estrutura é o elemento de inteligência que reduz a acentuada tendência entrópica de massas de dados.

Entre os principais cadastros de que o Sistema Estatístico Nacional carece deve-se citar: *a)* o cadastro das empresas, unidades de produção e estabelecimentos, os últimos consoante a definição do IBGE; *b)* o cadastro de domicílios; *c)* o cadastro da população.

Entre as principais tabelas, as de natureza funcional e espacial que mencionamos anteriormente: produtos, municípios, setores censitários, etc.

O Sistema Estatístico Nacional tem de esclarecer como distribuir a responsabilidade da custódia dos principais cadastros e tabelas nacionais entre os seus órgãos. O mentor dos padrões, formalizado institucionalmente, é o IBGE.<sup>14</sup>

A conseqüência principal da existência de alinhamento estrutural em bases de dados é a oportunidade de ampliação das associações de dados pelo exercício de *matching*. Cumpre agora revermos o conceito de *matching* e examinarmos a extensão de sua prática em outros países.

## 8. MATCHINGS

A associação de arquivos de diferentes agentes de um sistema estatístico, por *matching* de registros, exata ou estatisticamente comuns, permite considerável ampliação de bases de dados interagências.

O exemplo clássico é o modelo de Joseph Pechman, do Instituto Brookings, para a Secretaria da Receita Federal dos EUA,<sup>15</sup> em 1962.<sup>16</sup> Pechman simulou os efeitos de uma alteração tributária num arquivo fiscal de 100 mil registros selecionados, aos quais adicionou dados extraídos da previdência social, como idade, sexo e raça de contribuintes. O fisco forneceu à previdência um arquivo de números de registro social e esta lista, sem o reconhecimento nominal dos contribuintes, constituiu a ponte para um *match* exato entre os dados das organizações. O sucesso do modelo provocou sua pronta adoção como método de estimativa de receita tributária pelo Departamento do Tesouro Americano.

Um *match* é exato quando um código comum e inequívoco é contido em registros de arquivos distintos. Um *match* é estatístico quando a associação tem fundamento em algum critério de similitude entre registros, tornando a correspondência exata um evento provável mas incerto, e, sobretudo, desnecessário.

*Matchings* estatísticos estão na raiz da metodologia de inúmeras sínteses de dados; por exemplo, as contas nacionais, assim como na imputação de valores por meios econométricos a campos falhados.

*Matchings* estatísticos exigem arquivos densos e naturalmente dependem do alinhamento conceitual e estrutural entre variáveis de um e de outro arquivo. Alguns são muito precisos, como por exemplo os que decorrem da *fonetização* mecânica de nomes de pessoas, parentes, endereços, etc. Suprimindo vogais, partículas, brancos, caracteres especiais e uma dentre cada dupla consoante em nomes de pessoas, fabrica-se uma chave de associação de alta taxa de acerto entre arquivos, exercício que pode ser amplamente elaborado para distinguir homônios, pela concatenação de novos discriminantes, idade, sexo, consoantes do endereço, etc.

*Matchings* estatísticos são, principalmente, associações despidas de controvérsia sobre questões de sigilo e segurança pessoal, abrindo caminho para a criação de amplas bases de dados contendo informações demográficas tributárias, de previdência, emprego, saúde, habitação, etc., oriundas de diversas fontes.

As mais importantes aplicações de associação de arquivos, entretanto, parecem ter apoio em *matchings* exatos.<sup>17</sup> Por exemplo, os arquivos LEED (Longitudinal Employer-Employee Data) e CWHHS (Continuous Work History) da Previdência Social dos EUA; o 1973 Exact Match File associando dados censitários, de previdência e receita federal, em 50 mil domicílios daquele país; o emprego do registro nacional da população em variados estudos econômico-sociais dos países escandinavos; o estudo longitu-

dinal da renda de indivíduos 1971/75 no Canadá, unindo elementos do censo e do fisco, etc.

No Brasil, associações deste tipo devem ser episódicas, exigindo inusitado esforço e atendendo demanda muito especial, certamente fora do âmbito do usuário comum. Isto se deve a que, em nosso meio, a reconstrução do alinhamento formal e conceitual de dados é matéria que requer estudo prolongado, pelo fato de serem descomprometidos e divisionais os cadastros, tabelas e padrões existentes.

Entretanto, os exemplos de *match*, exatos e estatísticos, em sistemas estatísticos de países mais avançados revelam como tem sido comum o emprego das trocas de dados entre arquivos de organizações censitárias e administrativas naqueles países, em apoio da análise de uma ampla faixa de problemas. Transferências de dados entre o censo e o fisco nos EUA, sem quebra de sigilo pessoal, têm lugar há pelos menos 15 anos.<sup>18</sup>

O reconhecido avanço do sistema estatístico canadense, por sobre quase todos os demais, decorre, nesta linha de raciocínio, da antecipada visão que aquele país teve (indo além de fixar padrões comuns de arquivamento de dados) da vantagem das organizações de dados em disco magnético sob controle de supervisores de *software* especiais. Rapid, um supervisor desenvolvido pelo birô do censo canadense,<sup>19</sup> pauta-se em instrumentos, *software* e *hardware*, do conhecimento de instalações brasileiras há alguns anos. O avanço daqueles países sobre o Brasil, nesta matéria, é resultado de que despertaram para o problema das integrações de dados no momento certo, há mais de 15 anos, mais do que na dificuldade da assimilação das técnicas estatísticas e computacionais que empregam.

*Matchings* calcados em cadastros, tabelas e padrões dão nova dimensão à união dos repertórios de dados divisionais das agências estatísticas de nosso País. Como em outros países, entretanto, imediatamente reclamam a atenção sobre os perigos a que expõem o indivíduo, submetido que fica a um poder extraordinariamente ampliado da burocracia. Este é o tópico com que concluímos a exposição.

## 9. SIGILO E PRIVACIDADE

Nenhum assunto tem despertado tanta controvérsia, relativamente aos usos da computação eletrônica, como o conflito entre o indivíduo, sua segurança, privacidade e autodeterminação, e a supremacia dos sistemas mecânicos controlados por computador.

Com efeito, é impraticável arguir-se em favor de rigores disciplinares nas estruturas de dados de agências estatísticas, de transferências de dados interagências, de cadastros, tabelas e padrões, sem enfrentar diretamente a questão das fronteiras da ação do indivíduo, de um lado, e da associação temida entre o poder central, a burocracia e o computador, do outro.

Só a proposição de um cadastro da população, um dentre os necessários, já suscita o questionamento da

imposição e validade do identificador numérico único para as pessoas naturais, o qual, do ponto de vista estatístico, é a chave apropriada para o alargamento de bases de dados censitárias, previdenciais, fiscais, entre outras. A matéria é por demais vasta para merecer tratamento apropriado neste texto.<sup>20</sup> Sumário os principais aspectos da seguinte forma:

Sistemas de dados sobre pessoas jurídicas ou naturais são legal e socialmente aceitos sob as formas distintas de *inteligência* e de *informação estatística*.

Sistemas de inteligência, no sentido de vigilância, têm na observação do indivíduo o seu domínio de interesse e ocupam-se na construção de *dossiers* de comportamento do indivíduo em uma certa função, a exemplo das funções educacionais, de um educandário sobre seus alunos; médicas, de um clínico sobre seus pacientes; tributárias, da receita federal sobre seus contribuintes, etc.; mas também de funções controvertidas, como as de registro criminal e de segurança pública, que adicionam a informações factuais ingredientes imprecisos ou informes, e têm a propriedade de existirem em desconhecimento da pessoa investigada.

Sistemas de *informação estatística* não têm no indivíduo o seu objeto de investigação, mas no conjunto de que ele faz parte.

Aqueles sistemas visam o estudo do comportamento de agregados de pessoas (e, mais genericamente, de itens de observação estatística) relativamente a uma certa função e, por isso, não se ocupam de pessoas distintamente, exceto como células de informação primária, mesmo nas circunstâncias em que a pessoa (ou item) é escolhida como observação de um painel e presta-se a uma repetição da medida de seus atributos ao longo de um certo tempo.

Sistemas de informação estatística compenetraram-se em dar resposta a questões como as que formulamos no início deste artigo e, portanto, apóiam decisões que visam o controle de fenômenos sócio-econômicos na empresa e no Governo.

No Sistema Estatístico Nacional, um sistema de informação estatística, agentes e respondentes estão protegidos por lei, sendo vedado o uso de informações para fins de inteligência.<sup>21</sup>

Esta distinção entre sistemas estatísticos e de inteligência seria, de *per se*, tecnicamente suficiente para remover resistências quanto à adoção, pelo Sistema Estatístico Nacional, das medidas de padronização exigidas para o seu alargamento. Contudo, se não as remove (é explícita a oposição a um cadastro da população, em diferentes segmentos sociais) é pelo fato de o problema não residir dentro de fronteiras técnicas, exclusivamente, mas de apoiar-se também em fatores de ordem psicológica, política e ética. No relatório das Nações Unidas:<sup>22</sup>

“It is imperative that statistical uses to employ adequate safeguards of the confidentiality of information about individual respondents and that the nature of these safeguards be widely understood. Paradoxi-

cally this is a problem that may become easier to deal with as integration progresses since, barring outright subversion, the danger of disclosure is greater where control is fragmented and there are no uniform rules. But faith in inefficiency as a safeguard is widespread and it is behind much of the political opposition to statistical integration.”

O argumento, dos que se opõem a padrões comuns de arquivamento e trocas de dados, está em que a diferença entre sistemas estatísticos e de inteligência não é nítida contextualmente (a pessoa mais rica e a maior empresa não precisam ser identificadas pelo nome; o endereço de uma pessoa não é informação neutra se for de uma instituição mental, etc.) e em que sistemas estatísticos podem ser penetrados para fins de inteligência, por não serem *smart-proof*. O caminho da integração é uma estrada sem retorno.

Já a defesa de valores humanos depende de valores éticos e políticos, de onde não ser impossível que a integração pretendida em sistemas estatísticos, para o apoio às decisões numa época, seja subvertida num instrumento de dirigismo das pessoas em outra época. A resistência tem raízes na conservação do direito do anonimato como dimensão da privacidade e no protesto contra a transferência à burocracia do último poder que resta ao cidadão (“privacidade é poder”).

Quando, contudo, o debate tem os ânimos serenados, a poeira dos argumentos filosóficos, ideológicos e políticos assentada e os cenários mais fantásticos de dirigismo mecânico do cidadão colocados em perspectiva, a questão resolve-se de forma simples: a integração de dados elementares em arquivos estatísticos e administrativos não é objeto de uma escolha aberta à sociedade, mas de uma escolha que ela já tomou e que se encontra em curso de implantação, decorrência do impulso *biológico* das organizações humanas em compartilharem recursos comuns.<sup>23</sup>

Assim, nosso País não conta com um cadastro da população, peça de um sistema estatístico, mas tem um CPF, que se insinua como identificador universal. E quem examinar a abrangência de alguns sistemas administrativos atuais em nosso meio, com PIS-Pasep-Rais e Polvo<sup>24</sup> deve entender que o recurso da defesa do cidadão não está em obstruir a tecnologia mas em regulamentá-la.

O problema não reside no dado em si, mas em seu uso. O controle do uso de sistemas estatísticos para fins de inteligência, do emprego do computador em complementação de *dossiers* manuais, da mecanização de dados errados e incompletos e da sua transposição a outros arquivos, etc. é matéria que uma sociedade resolve fora do terreno da tecnologia, no domínio da lei.

Num mundo intensamente dirigido para a mecanização, uma pessoa, ao longo de sua vida, fornece novas e mais freqüentes informações e aceita participar de um processo que dela exige cada vez mais dados como parte das regras do jogo da convivência. A dimensão do indivíduo é algo a ser procurado fora do que contém as fitas magnéticas. Donald Michael:<sup>25</sup>

“Who is the real me? The ‘me’ in the databanks that not only include the endless raw data about me — assume it is all correct — but also that which is organized into models that permit sophisticated, economic, biological and social psychological representations of my behavior and analysis and interpretations thereof, models I may not even understand? Or am I what isn’t in the tapes? For some people the real me will be beyond the tapes; for others, ‘I’ am what is in the tapes because that is how I am recognized and responded to an increasingly credit-card, computer educated, social indicator society that rewards or punishes me for being what the databanks say I am. Some individuals will feel themselves both kinds of ‘me’s’. And since a crucial part of one’s sense of ‘me’ will be for every sane person what it always has been, namely what I think you think is ‘me’, questions will also burden some in the form who is the real you?”

O que somos ou seremos, você e eu, leitor, respondendo a Michael, depende em grande medida do equacionamento dos problemas econômicos e sociais, em nível de empresa e de Governo, no País. Evidentemente, é preciso contribuir para que as pessoas que decidem disponham de análises apoiadas em melhores dados e informações.

Claramente, entretanto, torna-se também necessário, a par das ampliações das bases de dados e da nova visão do detalhe, propugnada pelo público usuário do Sistema Estatístico Nacional, um exame de fundo na legislação brasileira, de forma a atualizar as relações desejadas por nossa sociedade entre a liberdade do indivíduo e o emprego dos sistemas estatísticos e de inteligência com os quais ele convive.

## 10. CONCLUSÃO

Os problemas do Sistema Estatístico Nacional dizem respeito, basicamente, à perda de foco sobre o seu objeto industrial, atualmente entendido como a síntese tabular impressa, situando o usuário numa posição de dependência na percepção que o sistema tem de suas necessidades. A atual situação é anacrônica, por não tomar em consideração que o aparelhamento e a competência de usuários têm de ser elevados para dar resposta, em sínteses próprias, à multiplicidade das questões de ordem econômica e social que o sistema não pode, por antecipação, fabricar. Parece em ordem o desejo do usuário atual de que o sistema lhe permita o acesso ao dado primário, imagem da coleta, em bases de dados consistidas e certificadas pelo sistema, sem quebra do sigilo da informação. Sendo descentralizado o mecanismo processador do Sistema Estatístico Nacional, parece importante apressar a capacidade do *matching* de arquivos inter e intra-agências pelo estabelecimento de cadastros, tabelas e padrões comuns de processamento. Estes instrumentos são indispensáveis para restaurar rigor conceitual e comparabilidade, a nível de microdados divisionais.

A demora do nosso sistema estatístico, em alinhar-se com os sistemas estatísticos de outros países naquele propósito, não é decorrência de falta de

assimilação, pelo meio nacional, de técnicas estatísticas ou de processamento de dados.

Mais provavelmente é consequência da falta de um ponto de referência técnico, normativamente predominante, no sistema; do embate das culturas técnicas e administrativa nas agências do sistema; e, mais amplamente, da situação cultural do País, onde gerações novas e antigas esforçam-se numa convivência dificultada pela aceleração dos eventos tecnológicos e sociais dos últimos vinte anos.

Não se deve recear que a disputa entre a liberdade do indivíduo e a inércia do binômio burocracia-computador não possa ser resolvida dentro da lei, nem se deve impedir a sociedade de colher os melhores benefícios desses novos mecanismos, para o aperfeiçoamento das decisões da empresa e do Governo, por adesão à ineficiência como forma de preservar dimensões supostamente seguras de privacidade. Mais pragmático é entender-se que sistemas mecânicos têm, em sua integração, uma forma natural, *biológica*, de enriquecimento que deve ser regulamentada com inteligência.

Dentro do Sistema Estatístico Nacional, a tendência à ação integradora de microdados, inter e intra-agências é observada em inúmeros exemplos dos últimos 10 anos. Tudo leva a crer que as presentes dificuldades por que passa o sistema estatístico são transitórias, em face do progresso do País, e que o seu amadurecimento até o nível das associações de dados exigidas em países mais complexos e avançados poderá ser atingido em futuro próximo.

34

<sup>1</sup> O Sistema Estatístico Nacional, previsto na Constituição, compreende, pela Lei nº 6.183, de 11 de dezembro de 1974, as estatísticas primárias, as estatísticas derivadas e a sistematização de dados sobre o meio-ambiente. A lista destes elementos é esmiuçada no Decreto nº 74.084, de 20 de maio de 1974. Integram o sistema todos os órgãos governamentais de administração direta e indireta, em todos os níveis, e os privados que exerçam atividades estatísticas com subvenção ou auxílio de órgãos públicos.

<sup>2</sup> Kaysen, C. et alii. *Report of the task force on the storage of and access to government statistics*. Washington D. C., Bureau of the Budget, Oct. 1966.

<sup>3</sup> *The Computer and invasion of privacy*. Hearings Before a Subcommittee of the Committee on Government Operations, House of Representatives, U. S. Government Printing Office, July 1966.

<sup>4</sup> Cumpre reconhecer o empenho do sistema estatístico, nos últimos 20 anos, em equiparar-se, consolidar o seu programa de estatísticas contínuas e dar características multidisciplinares a algumas de suas bases de dados. Os principais exemplos de estatísticas contínuas no País são a Pesquisa Nacional de Amostragem de Domicílios (PNAD) e as Pesquisas Industriais (PI), de regime mensal e anual, séries que tiveram origem em décadas passadas. Todas adquiriram rigor e abrangência em anos mais recentes. A série das PNADs foi interrompida em 1974/75 para a realização de um inquérito de singular fôlego e implicações multidisciplinares: o Estudo Nacional de Despesa Familiar (Endef), que inovou até em pro-

cedimentos de armazenamento de dados em computador. Conquanto estes e outros progressos devam ser creditados ao sistema, não há como escapar à conclusão de que resta ainda muito por fazer.

<sup>5</sup> Kerstenetzky, I. O Papel da informática no desenvolvimento. In: Silva, Benedicto et alii. *Da documentação à informática*. Rio de Janeiro, Fundação Getúlio Vargas, 1974.

<sup>6</sup> Existem sínteses que estão fora do alcance do usuário comum, como, por exemplo, as contas nacionais, matriz de insumo-produto, matriz de energia, etc., razão de o sistema estatístico incluir em seu propósito as chamadas estatísticas derivadas. O usuário perde, contudo, ainda assim, condição de análise se não dispuser de acesso a arquivos selecionados empregados naquelas sínteses.

<sup>7</sup> Ministério de Minas e Energia. Portaria nº 958, de 6 de agosto de 1976.

<sup>8</sup> Churchman, C. W. & Schainblatt, A. H. The Researcher and the manager: a dialectic of implementation. *Management Science*, Feb. 1965.

<sup>9</sup> Argyris, C. Management information systems: the challenge to rationality and emotionality. *Management Science*, Feb. 1971.

<sup>10</sup> Simon, H. *Administrative behavior*. Mac Millan, 1961.

<sup>11</sup> Simon define autoridade como o poder de decidir, de forma a orientar a ação de outros. Nestes termos, a relação entre superior e subordinado é intercambiável, conforme o objeto da decisão. Obediência, como abdicação de escolha, é decorrência de autoridade conferida em outro sentido, no sentido administrativo do direito à última palavra. Veja Simon, H. op. cit.

<sup>12</sup> *The Computer and invasion of privacy*. op. cit.

<sup>13</sup> Uma base de dados contém os dados puros, imagem da medida, desprovidos de ponteiros e organizações, introduzidas por *software* supervisor. Uma base de dados supervisionada por um Data Base Management System (DBMS) é uma construção mais organizada, talvez um banco de dados.

<sup>14</sup> Art. 5º da Lei nº 6.183, de 11 de dezembro de 1974, que instituiu o Sistema Estatístico Nacional.

<sup>15</sup> *The Computer and invasion of privacy*. op. cit.

<sup>16</sup> *Methods of collecting, organizing and retrieving social statistics to achieve integration*. Report of the Secretary General, United Nations Economic and Social Council, E/CN 3/516, June 1978.

<sup>17</sup> *Methods of collecting, organizing...* op. cit.

<sup>18</sup> Bowman, R. Depoimento. In: *The Computer and invasion of privacy*. op. cit.

<sup>19</sup> Turner, M. J.; Hammond, R. & Colton, P. *DBMS for large statistical databases*. In: *Proceedings of Fifth International Conference Very Large Databases*. Rio de Janeiro, out. 1979.

<sup>20</sup> Westin, A. *Privacy and freedom*. Atheneum, 1970; Westin, A. & Baker, M. *Databanks in a free society, report to the National Academy of Sciences*. Quadrangle Books, 1972; *Privacy and computers*. Report of a Task Force, Department of Communications/Department of Justice, Information Canada, Oct. 1972; Armer, P. The Individual: his privacy, self image and obsolescence. *The Management of information and knowledge*. Committee on Science and Astronautics, U. S. House of Representatives, 1970.

<sup>21</sup> Art. 1º, Lei nº 5.534, de 14 de novembro de 1968. O artigo não menciona *inteligência*, mas é suficiente em estabelecer o uso de dados "exclusivamente para fins estatísticos".

<sup>22</sup> *Methods of collecting...* op. cit.

<sup>23</sup> Baran, P. Depoimento. *The Computer and invasion of privacy*. op. cit.

<sup>24</sup> Projeto Polvo: cadastrar, cadastrar. *O Globo*, 7 maio 1978.

<sup>25</sup> Michael, D. The Individual: enriched or impoverished? Master or servant? *Information technology: some critical implications for decision makers*. The Conference Board, 1972.