

REDATAM: UMA FERRAMENTA  
PARA OBTER INFORMAÇÃO DESAGREGADA GEOGRAFICAMENTE  
PARA O PLANEJAMENTO

Ari N. Silva

Documento a ser apresentado no Seminário Latinoamericano de Compilação e Procesamento de Dados Demográficos, Santiago de Chile, 23 a 27 de maio de 1988 (original em espanhol).

COMISSÃO ECONOMICA PARA A AMÉRICA LATINA E CARIBE (CEPAL)  
CENTRO LATINOAMERICANO DE DEMOGRAFIA (CELADE)  
Casilla 91 - Santiago de Chile  
Tel: 2283206 Cable: UNATIONS

CELADE

DOCUMENTO  
MICROFILMADO

CEPAL

## INDICE

I. INTRODUÇÃO . . . . .	1
II. O SISTEMA REDATAM . . . . .	3
III. RESUMO DAS FACILIDADES DO SISTEMA REDATAM . . . . .	5
IV. LIMITAÇÕES DO SISTEMA . . . . .	8
V. EQUIPAMENTO NECESSARIO PARA USAR O REDATAM. . . . .	9
VI. EXPERIENCIAS NA UTILIZAÇÃO DO REDATAM . . . . .	10
VII. REDATAM PARA O CENSO DE 1980 DO BRASIL . . . . .	13
VIII. REDATAM NO FUTURO . . . . .	14
IX. CONCLUSOES . . . . .	16
APENDICE: NECESSIDADES PARA A CRIAÇÃO DE UMA BASE DE DADOS REDATAM . . . . .	18
1. Equipamentos . . . . .	18
2. Características dos dados . . . . .	18
3. Pessoal . . . . .	18
4. Tempo . . . . .	19
5. Espaço ocupado e tipo de geração . . . . .	19

## I. INTRODUÇÃO.

1. O objetivo deste documento é apresentar a experiência adquirida pelo CELADE na utilização da informação censitária através de um sistema computacional chamado REDATAM 1/, que permite explorar os dados especialmente num nível de áreas geográficas pequenas, tais como cidades, bairros ou quarteirões, ou qualquer agrupamento destas, a um custo reduzido e diretamente pelo usuário, por meio da utilização dos microcomputadores.

2. Em junho de 1983, um técnico do Centro Latinoamericano de Demografia (CELADE), visitou seis países da América Latina e do Caribe -Brasil, Costa Rica, Chile, Peru e Trinidad e Tobago- com o propósito de examinar a situação dos usuários de dados estatísticos. 2/

3. Como resultado desta missão, se comprovou a existência de sérias dificuldades em termos de disponibilidade da informação sobre população, que se descrevem em seguida: 3/

a) O desenvolvimento de novos projetos ou o melhoramento dos serviços sociais têm uma dimensão espacial que se reflete na localização dos mesmos. Dentro dos muitos fatores que se devem considerar ao efetuar o planejamento de tais projetos, estão as características e a distribuição geográfica da população que será beneficiada e/ou afetada pelos mesmos.

b) Os dados de população de uso mais frequente no país provêm dos censos de população e domicílio, mas, em geral, as áreas geográficas do projeto dificilmente correspondem aos limites político-administrativos; isto faz com que o usuário tenha a necessidade de "construir" sua área de interesse através de uma composição das subdivisões geográficas, tais como os setores censitários.

c) É praticamente impossível que os institutos de estatística

---

1/ REDATAM = REcuperação de DADOS de Areas pequenas por Microcomputador

2/ Conning, Arthur M., Report to IDRC of the REDATAM Pre-Project Mission, CELADE, Santiago, Chile, outubro de 1983.

3/ Conning, Arthur M., Geographically Disaggregated Census Data for Planning in Developing Countries, International Union for the Scientific Study of Population Conference, Florencia, Italia, 1985.

publiquem (ou armazenem os relatórios do computador) a totalidade da informação sobre as áreas menores do país. Supondo-se que um usuário requeira um conjunto de quadros de uma área específica que não corresponda a nenhuma das tabulações existentes (ou a uma combinação delas), a única solução é reprocessar os dados censitários de acordo com as necessidades específicas.

d) A maioria dos institutos de estatística da região não está equipada para reprocessar os dados censitários em forma rápida e com um custo baixo; primeiro, porque os dados são volumosos e estão organizados convencionalmente (em fitas magnéticas). Segundo, uma vez terminado o censo, tanto computadores como programadores têm seu tempo dedicado a outros serviços com maior prioridade. As solicitações de processamentos especiais para áreas pequenas devem, geralmente, esperar semanas ou meses e às vezes têm um custo total muito grande.

4. Existe um tipo de usuário em potencial que resulta sistematicamente prejudicado pelas quatro características mencionadas: são aqueles que requerem informação sobre áreas geográficas pequenas, tais como os departamentos municipais de planejamento, as empresas particulares, as companhias prestadoras de serviços, etc. Quase sempre se trata de projetos com escassos recursos, sem possibilidades de poder enfrentar os custos de processamento do arquivo do censo, com pouca ou nenhuma relevância em termos nacionais, e suas demandas são insatisfeitas.

5. A tendência do mercado de usuários do censo é de multiplicar sua quantidade ao diminuir o universo de estudo. A ênfase atual está na descentralização e na explosão, ou pulverização dos atores nos níveis menores do processo de planejamento.

## II. O SISTEMA REDATAM.

6. Com o objetivo de solucionar os problemas enumerados, e diante da ausência de opções no mercado de "software", se decidiu construir um sistema especial ad hoc. 4/ Para uma maior flexibilidade de sua utilização, se optou por armazenar todos os microdados censitários (isto é, o valor direto de cada variável - sexo, idade, ocupação, etc. - para cada indivíduo ou registro do censo). A grande vantagem desta estratégia é a possibilidade de elaborar qualquer quadro, independente do nível de detalhe, para a área ou grupo de áreas que se solicite. A possível desvantagem seria o tamanho desta base de dados, que poderia ser superada com uma cuidadosa técnica de compressão na etapa de armazenamento.

7. Tomando-se em conta as limitações (tempo e custo) do acesso aos grandes computadores e a seus programadores, uma componente central da solução proposta está baseada no uso dos microcomputadores, de baixo custo e de uso corrente (IBM ou compatíveis), eliminando-se assim a dependência dos usuários aos recursos computacionais centrais dos institutos de estatística. Estes equipamentos se tornam cada vez mais baratos, estão disponíveis na região e podem ser usados diretamente pelos próprios investigadores e por outras pessoas que não necessariamente sejam programadores.

8. A filosofia embutida no REDATAM se reflete no espaço necessário para armazenar uma base de dados, que é aproximadamente quatro vezes menor que o espaço inicial ocupado pelos arquivos em fita magnética, e no tempo de processamento das solicitações dos usuários. Por exemplo, se se requer uma tabela de idade por sexo, não é necessário ler todas as variáveis da base. Somente se necessita processar as variáveis envolvidas, o que aumenta consideravelmente a velocidade de execução. Além disso, ao utilizar-se os ponteiros para as áreas geográficas menores, não é necessário recorrer todo o arquivo para leer, por exemplo, os dados de um conjunto de quarteirões.

---

4/ Com fundos do Centro Internacional de Investigações para o Desenvolvimento (CIID) do Canadá, e o apoio adicional do Fundo das Nações Unidas para Atividades em Matéria de População (FNUAP) e da Agência Canadense para o Desenvolvimento Internacional (ACDI).

9. Não se pode comparar a velocidade dos microcomputadores com a dos equipamentos maiores, nos quais um proceso pode durar duas horas, sendo que nos últimos talvez só demoraria cinco minutos. Porém, a dificuldade de acesso dos usuários aos equipamentos grandes e a dependência aos programadores, faz com que seja preferível utilizar o microcomputador, que está ao alcance da mão, diante da outra alternativa de esperar dias ou semanas para que a máquina grande esteja disponível. Por outra parte, na grande maioria das solicitações de dados sobre áreas pequenas, não se necessita mais que alguns minutos para identificar a área de interêsse e obter os resultados.

10. Uma vantagem adicional do sistema REDATAM é sua capacidade de processamento hierárquico, que permite obter resultados cruzados entre variáveis de domicílio e população como, por exemplo, o número de estudantes por sexo y tamanho do domicílio que vivem sob determinadas condições referidas ao estado do teto e das paredes da casa.

11. A versão 1.00 do sistema REDATAM foi instalada nos dois países piloto (Chile e Santa Lúcia durante o ano de 1986). A versão 2.00 começou a distribuir-se em abril de 1987 e atualmente está disponível a versão 3.00, recém terminada em março de 1988 (em fase de documentação final). Existem três manuais de utilização do sistema 5/, 6/ e 7/, e tanto o sistema como os manuais se encontram disponíveis em Espanhol e em Inglês (existem em andamento contatos com o IBGE para a tradução dos documentos e do sistema para Português).

---

5/ CELADE, REDATAM, Manual do Usuário, (LC/DEM/G.50), Santiago, maio de 1987.

6/ CELADE, REDATAM, Manual de Geração de Bases de Dados, (LC/DEM/G.53), Santiago, outubro de 1987.

7/ CELADE, REDATAM, Facilidades da Nova Versão 3.00, a ser publicado.

### III. RESUMO DAS FACILIDADES DO SISTEMA.

#### 12. Seleção geográfica

Definição do universo de casos para processamento, selecionando-se somente as áreas geográficas pertinentes.

#### 13. Agrupação de áreas geográficas

Possibilidade de compor conjuntos geográficos de interesse a partir de áreas pequenas, ainda que estas não estejam sob a mesma hierarquia geográfica, como no caso das áreas metropolitanas, ou o estudo de municípios de dois estados distintos mas que utilizam uma mesma estrada (ou banhados por um mesmo rio).

#### 14. Base de dados autodocumentada

As variáveis têm um nome curto (usado na programação); um título (para imprimir com os resultados); definição das categorias existentes (com códigos e nomes) e outros parâmetros técnicos.

#### 15. Utilização interactiva ou em forma encadeada

Funciona com base em menus e com muita ajuda direta ao usuário. Também deteta muitos erros de digitação de comandos.

Se pode também pré-programar vários processos demorados e executar-los de uma forma encadeada, sem a intervenção humana, e num horário adequado (durante a noite, por exemplo).

#### 16. Cálculo de variáveis derivadas

Utilização de operações aritméticas e instruções de recodificação para obtenção de novas variáveis, como grupo quinquenal de idade, ou o índice de qualidade do domicílio a partir da estrutura do teto, piso e paredes. Estas podem ser temporárias (durante o processo) ou incorporadas à base de dados.

#### 17. Resultados estatísticos

Produção de três tipos de estatísticas: frequências, tabulações cruzadas e médias, estes dois últimos com até quatro variáveis. É possível obter

resultados com porcentagens em relação ao total de linhas, de colunas ou ao total geral, assim como resultados para o total da área de interesse e/ou para os níveis geográficos intermediários.

**18. Procesamento hierárquico**

Existência de dois níveis de variáveis (domicílio e população), com obtenção de resultados para cada um deles, ou para ambos conjugados. Por exemplo, o número de pessoas por sexo e grupos de idades que vivem sob determinadas condições de teto e paredes.

**19. Geração de sub-bases**

Facilidade para criar pequenas bases de dados REDATAM que incluam somente as áreas de interesse de um determinado estudo. A partir da base a nível nacional, se podem criar, por exemplo, bases de dados por estado e estas serem entregues aos centros regionais.

**20. Utilização de fatores de ponderação (pesos)**

Aplicação de pesos aos resultados estatísticos com o fim de obter totais ponderados como nos censos por amostra ou nas pesquisas de domicílio.

**21. Produção de arquivos para outros programas**

Supondo-se que as operações estatísticas básicas não sejam suficientes, se pode gerar arquivos com as variáveis de interesse e com as observações correspondentes à área selecionada, para serem processados posteriormente por outro programa como o SPSS, o SL-MICRO, ou um programa escrito pelo usuário.

**22. Armazenamento de resultados**

Se podem guardar as saídas dos processos estatísticos para uma posterior impressão, sua incorporação em textos de documentos, assim como para a produção de gráficos, usando-se outros programas tais como Wordperfect, Wordstar, Lotus, Mathplan, etc.

**23. Criação de bases de dados em forma descentralizada**

Possibilidade de criar bases de dados parciais (regiões, por exemplo), e de conectar-las no fim, facilitando o processo no caso de arquivos volumosos.

**24. Proteção à informação**

Implementação dos conceitos de segredo estatístico da informação e proteção dos dados da base, incluindo "passwords" ("chaves") para as funções de administração e para o acesso aos distintos níveis geográficos, com o objetivo de respeitar a confidencialidade e impedir a identificação de domicílios e pessoas por usuários não autorizados. Proteção também para a criação de sub-bases de dados, com a eliminação, por eleição, dos níveis mais baixos da hierarquia geográfica, possibilitando assim a exportação de sub-bases de dados com a devida proteção da confidencialidade.

**25. Variáveis decimais**

Implementação do armazenamento e cálculo de variáveis decimais.

#### IV. SUAS LIMITAÇÕES.

##### 26. Equipamentos IBM

O sistema foi desenhado para ser operado em equipamentos IBM/PC/XT/AT ou compatíveis. Já foi provado seu funcionamento em máquinas EPSON, TOSHIBA, COMPAQ e GREAT WALL (computador chinês).

##### 27. Somente dois níveis hierárquicos

Se o censo tem três níveis (domicílio, família e pessoas), se gera família e pessoas como primeiro e segundo níveis respectivamente e, para a obtenção de resultados a nível de domicílio, se selecionam somente as primeiras famílias. Para outro tipo de arquivos com mais de dois níveis não há solução.

##### 28. Tempo de procesamento para seleções geográficas maiores

Não se pode comparar a velocidade dos microcomputadores com a das máquinas grandes ("mainframe"). Para um país ou região muito grande em termos populacionais, e para um estudo no qual o universo de interesse é todo o país (ou região), talvez seja preferível utilizar o equipamento central.

##### 29. A criação de uma base de dados nem sempre é uma tarefa fácil

Para censos e pesquisas relativamente pequenos nos quais é factível o armazenamento inicial dos dados em forma convencional num microcomputador, a criação da base REDATAM é relativamente fácil. Porém, para os arquivos maiores, o processo resulta complexo e se necessita de um programador com conhecimentos do equipamento central e de microcomputador. Em ambos casos, é imprescindível a existência de uma linha de comunicação entre os dois equipamentos.

##### 30. Falta de integração com uma base de informação cartográfica

A seleção geográfica tem que ser feita com base nos códigos e nomes das localidades (estados, microrregiões, municípios, distritos, etc.), o que exige uma prévia consulta de mapas para definir o conjunto de códigos a ser utilizado.

## V. EQUIPAMENTO NECESSARIO PARA USAR O REDATAM.

31. IBM PC, XT, AT, ou compatível.
32. Memória interna de 640 Kb.
33. Uma unidade de disco flexível.
34. Um monitor (terminal) monocromático ou colorido.
35. Uma impressora com um carro de pelo menos 80 caracteres de largura.
36. Sistema operacional PC-DOS versão 2.0 ou mais atualizada.
37. Um disco duro com aproximadamente 1,4 Mb disponível para o sistema REDATAM e base de dados de demonstração.
38. O espaço necessário para o armazenamento de um censo ou pesquisa de um país dado dependerá do número de pessoas recenseadas e do número de variáveis do questionário (ver apêndice).

## VI. EXPERIENCIAS EM SUA UTILIZAÇÃO.

39. O sistema REDATAM tem sido utilizado em vários países. A tabela 1, a seguir, detalha os tamanhos e espaços de armazenamento requeridos em cada caso, e à continuação se descrevem brevemente os detalhes de cada um dos países, 8/ y 9/, com alguns comentários segundo sua necessidade.

Tabela 1: Armazenamento requerido para bases de dados REDATAM

País	Domicílios (mil)	Pessoas (mil)	Espaço (Megabytes)	Meio de armazenamento
Chile	4.000	12.000	300	Disco Laser
Santa Lúcia	30	125	3	Disco Duro
Costa Rica	500	2.500	60	Disco Duro
Guiana (pesquisa)	8	42	1	Disco Duro
Uruguai (amostra)	75	450	10	Disco Duro
Dominica	17	74	2,6	Disco Duro
Colômbia (básico)	5.800	27.800	70	Disco Duro
Colômbia (ampliado)	600	2.800	60	Disco Duro

40. Na tabela acima não aparecem as bases de dados geradas para os países caribenhos de Granada, Ilhas Virgens e St. Vincent, as quais ainda não foram instaladas.

### CHILE

Duas bases de dados, uma para o censo completo de 1982, e outra para uma amostra de 5 por cento (cêrca de 600 mil pessoas).

Estudo de domicílios coletivos da Grande Valparaíso, a nível de comunidade, distrito y zona.

Estudo da população da cidade de Concepción, a nível de distritos e zonas censitárias, por condição de ocupação, atividade, grupos de idade e

---

8/ Silva, Ari, REDATAM: Relevância para os Censos de 1990, (LC/DEM/R.48), CELADE, Santiago, Chile, setembro de 1987, 9 pp.

9/ Silva, Ari, REDATAM: Um Resumo, (LC/DEM/R.50), CELADE, Santiago, Chile, setembro de 1987, 12 pp.

nível de instrução.

Aplicação ao projeto da rede de esgoto para a cidade de La Serena, a nível de distritos, zonas e quarteirões.

Aplicação ao projeto de resultados especiais para três comunidades de Santiago, com tabulações a nível de bairros e criação de sub-bases de dados para cada uma das comunidades.

Estudo das características da população indígena da província de Temuco.

#### COSTA RICA

Base de dados do censo de 1985.

Projeto de Desenvolvimento dos municípios de Guatuso e Upala.

Projeto de Saúde para o estudo do envelhecimento da população no município de Puriscal.

Projeto para os centros regionais das Universidades, em Turrialba e Nicoya.

#### GUIANA

Base de dados da Pesquisa Demográfica Retrospectiva GUYREDEM de 1986.

Análise dos resultados da pesquisa.

Tabulações de fecundidade para as nove regiões do país.

#### SANTA LUCIA

Base de dados do censo de 1980.

Obtenção de resultados a nível de país para pedidos de diferentes usuários (Ministério da Saúde, Agricultura, etc.).

Estudo da expansão telefônica para alguns bairros de Castries.

Estudo para a construção de uma represa.

Aplicação al projeto de expansão da rede de esgotos.

#### URUGUAI

Base de dados da amostra de 15 por cento para resultados antecipados do censo de 1985.

Projeto para o estudo da pobreza crítica de todas as regiões do país, incluindo características do domicílio e da população.

Tabulações de verificação para a publicação de resultados avançados da

amostra.

41. No caso de Santa Lúcia, um país que não dispõe de um computador grande e cujo censo foi processado em Barbados, o sistema REDATAM é utilizado para TODAS as solicitações de tabulados.

42. Na Guiana, o sistema se usou para carregar os dados da Pesquisa Demográfica Retrospectiva (GUYREDEM, Retrospective Demographic Survey) de 1986, que apesar de não ser muito significativa em termos de volume de informação, serviu para provar o sistema com arquivos ponderados (fator de expansão).

43. No caso de Costa Rica, é importante comentar que alguns usuários externos à Direção Geral de Estatística e Censos (por exemplo, a Companhia de Luz), se mostraram muito interessados em ter um subconjunto de variáveis da base de dados do censo que pudesse ser processada em seus próprios microcomputadores. Este projeto está sendo estudado na Direção de Estatística. Também se gerou uma base de dados para a Pesquisa de Domicílios de 1986 e ainda que os fatores de expansão não permitiam que se tivessem resultados de áreas muito pequenas por tratar-se de uma pesquisa por amostragem, foi possível processar um cruzamento de variáveis para toda a pesquisa em menos de cinco minutos. Outro arquivo criado em REDATAM foi o do sistema de Registro Civil com os nascimentos, casamentos e mortes.

44. Como no caso de Colômbia o censo de 1985 foi realizado por amostra, se geraram duas bases de dados, uma para o universo do país e outra para o questionário ampliado, a segunda com a restrição de que não se podem obter resultados de áreas muito pequenas porque os fatores de expansão não são confiáveis nestes níveis. Um detalhe interessante no caso colombiano é que na verdade se geraram seis bases de dados regionais, de maneira que possam ser usadas em forma descentralizada nas delegacias regionais do Departamento Administrativo Nacional de Estatística (DANE).

## VII. REDATAM PARA O CENSO DE 1980 DO BRASIL.

45. O censo demográfico de 1980 do Brasil foi executado com dois questionários, o da amostra para 25 por cento da população, e o da não-amostra, para os 75 por cento restantes. O questionário da amostra tem 26 variáveis de domicílio e 61 variáveis de população, enquanto que o questionário da não-amostra tem somente 3 variáveis de domicílio e 7 de população.

46. Essa separação significa que deveremos ter duas bases de dados REDATAM, uma para a amostra (com todas as variáveis), e outra para o universo (só com as variáveis da não-amostra). Em termos de espaço ocupado para as bases de dados, baseando-nos num cálculo preliminar executado no CELADE, estimamos que, para cada milhão de pessoas numa determinada área (estado, município, etc.), são necessários: a) 17 Mb de disco para armazenar os dados da amostra e 4 Mb para os dados do universo.

47. Seria importante ter as duas bases em funcionamento porque, de acôrdo com a metodologia desta amostra de 25 por cento, ela não é suficientemente grande para garantir a obtenção de informações nos níveis mais baixos da hierarquia geográfica, isto é, não se pode pensar em utilizar-la para distritos ou setores censitários.

## VIII. REDATAM NO FUTURO.

48. Os técnicos do Instituto Nacional de Estatística, Geografia e Informática (INEGI) do México estão experimentando a geração de uma base de dados REDATAM para o censo econômico.

49. Existem convênios em marcha para a geração de bases de dados censitários para varios países da região como Bolívia, Cuba, Trinidad e Tobago e Venezuela. Destes o mais concreto é Trinidad e Tobago, onde se pretende gerar a base de dados neste semestre.

50. As autoridades brasileiras do IBGE envolvidas no censo de 1990, estão considerando a utilização do REDATAM para a obtenção rápida de resultados do censo piloto a ser realizado em setembro de 1988. Além disso, têm também a intenção de fazer uma distribuição maciça dos dados do censo de 1990 para uso público em forma de bases de dados REDATAM, que seriam enviados junto com uma cópia do sistema, permitindo aos usuários trabalhar diretamente com os dados em seus próprios microcomputadores.

51. O CELADE apresentou al CIID do Canadá um projeto para desenvolver, dentro de um período de dois anos, a segunda etapa do projeto REDATAM, a ser conhecido como REDATAM+. O novo sistema tem como objetivo solucionar as necessidades de um "banco de dados para o planejamento", capaz de armazenar e recuperar a informação multidisciplinária associada com dados de população e domicílio, com facilidades de manipulação cartográfica e visualização espacial da informação, particularmente nos níveis regionais e locais. Também o REDATAM+ está dirigido a garantir que os censos de 1990 da América Latina e do Caribe sejam capazes de prover a informação necessária para estes propósitos. O projeto foi aprovado e sua atividades iniciais começam em abril de 1988.

52. Estes objetivos refletem uma tendência geral na região, que é a de concentrar-se no planejamento setorial local e orientado espacialmente, com a visão de que os enormes gastos requeridos para a execução de um censo tenham um melhor retorno e uma maior utilização de sua informação que no passado.

53. Os grandes objetivos do REDATAM+ são os seguintes: a) Relacionar dados sobre população e outros temas com a informação cartográfica através da criação de um Sistema de Informação Geográfica (conhecido como GIS em Inglês); b) Associar um tipo de informação "ecológica" multidisciplinária sobre a descrição de áreas geográficas com dados de população e domicílio; c) Operar com a mesma base de dados numa rede de informação local ("Local Area Network"); e d) Produzir tabulações e resultados diretos para impressão ("camera ready").

54. O que fará do REDATAM+ uma ferramenta distinta das demais é que a informação "ecológica" 10/ que descreve as áreas geográficas podem ser relacionadas, de acordo com as necessidades do usuário, com o microdado "individual" de população e domicílio, que por sua vez podem ser processados com as especificações do usuário (na verdade, os dados ecológicos, poderiam haver sido gerados a partir dos dados individuais, através de agregações nos níveis geográficos superiores).

---

10/ Para os propósitos aqui, "dados ecológicos" são as variáveis que descrevem uma área geográfica, tais como o salário médio de um município, ou se um quarteirão conta com uma clínica de saúde. Por outro lado, dados "individuais" são as variáveis que descrevem pessoas, domicílios e outras unidades que são de interesse somente quando agregadas estatisticamente.

## IX. CONCLUSOES.

55. Resumidamente, o sistema REDATAM tem como propósito "organizar e manter os arquivos censitários, geograficamente classificados, dentro da capacidade limitada de um microcomputador, de tal modo que se possam gerar tabulações ou outras estatísticas para a unidade geográfica menor, como por exemplo bairro ou quarteirão, ou para qualquer agrupação destas". Assim, os arquivos do censo que antes somente podiam ser usados em grandes computadores, atualmente podem ser armazenados e processados em microcomputadores ao alcance do usuário.

56. Desta maneira, não existe a intenção de competitividade com outros sistemas ou produtos existentes no mercado para processamento das diversas etapas de censos e pesquisas. O REDATAM NÃO é um programa de "processamento" no sentido de "produzir" ou "transformar" a informação dos questionários. NÃO é para a entrada de dados, consistência nem tampouco para a produção de tabulados "camera ready". Seu ambiente, tal como já se descreveu no parágrafo anterior, pode ser traduzido mais como uma ferramenta de armazenamento de grandes volumes de informação, de fácil e rápido acesso direto pelo usuário final da informação, e a um baixo custo.

57. No caso dos censos da década de 90, ainda que não se utilize o sistema REDATAM frente à eventualidade de que apareçam ferramentas mais modernas, os conceitos e a experiência adquirida na sua utilização deveriam ser incorporadas às decisões sobre planejamento e organização, que a longo prazo afetam a posterior utilização dos dados para áreas pequenas.

58. Apesar de estar dirigido para a obtenção de resultados para áreas geográficas pequenas, o REDATAM pode ser de grande valia na análise de pesquisas por amostragem, sempre que sejam respeitados os limites impostos pela metodologia da amostra. Quer dizer, as áreas selecionadas para o processo devem conter um número mínimo de casos para que o resultado seja significativo.

59. Finalmente, cabe destacar que o sistema REDATAM, uma vez criada a base de dados, pode (e deve) prestar serviços ANTES que as publicações do censo

estejam prontas. Por outra parte, a facilidade e rapidez na obtenção da informação para qualquer nível geográfico, pode ter um impacto positivo na quantidade de quadros a serem publicados, limitando-os aos resultados mais gerais. Ao mesmo tempo, existe a possibilidade real de difundir a informação censitária a nível regional (descentralização da informação), mediante a geração de sub-bases REDATAM, com dados sobre a região de interesse e instaladas em equipamentos microcomputarizados, localizados em forma descentralizada.

## APENDICE: NECESSIDADES PARA A CRIAÇÃO DE UMA BASE DE DADOS REDATAM.

Através de exemplos, o manual de geração de bases de dados especifica as etapas de criação de uma base de dados REDATAM. Porém, como tópicos gerais, se pode mencionar o seguinte:

### 1. Equipamentos.

- 1.1. IBM/PC ou "full-compatible", de preferência um AT por razões de velocidade de processamento, com 640 Kilobytes de memória;
- 1.2. Disco duro para a criação dos arquivos e futuro armazenamento da base, dependendo do tamanho dos arquivos a serem gerados;
- 1.3. Disco laser (opcional), dependendo do tamanho dos arquivos a serem gerados para a base;
- 1.4. Comunicação entre o computador central e o micro para a transferência de dados;

### 2. Características dos dados.

- 2.1. Definição das variáveis geográficas e classificação do arquivo nesta mesma ordem;
- 2.2. Definição do dicionário com as variáveis do arquivo e obtenção de frequências das mesmas;
- 2.3. Cálculo do espaço definitivo a ocupar;
- 2.4. Verificação da qualidade das variáveis geográficas, com o objetivo de evitar problemas na geração dos índices;

### 3. Pessoal.

- 3.1. Quando o arquivo a ser criado é relativamente pequeno na sua forma original, ele pode ser armazenado diretamente no micro. Neste caso se pode recorrer ao procedimento de Carga Automática, para o qual se necessita só de uma pessoa com algum conhecimento de microcomputadores;
- 3.2. Quando o arquivo é grande, se necessita de um programador com alguma experiência em COBOL de computadores grandes, além de conhecimentos de microcomputadores;

#### 4. Tempo.

Em geral, em condições normais, uma base de dados pode ser criada num tempo que varia entre uma e três semanas, dependendo dos recursos disponíveis.

#### 5. Espaço ocupado e tipo de geração.

5.1. O espaço ocupado por uma base de dados REDATAM é função de quatro parâmetros básicos: o primeiro é o número de domicílios; outro é o número de variáveis perguntadas ao domicílio e o tamanho de cada uma delas. Ao aplicar-se o mesmo conceito também para os dados de população, se obtém os quatro parâmetros mencionados.

5.2. De acôrdo com a experiencia anterior, se pode calcular aproximadamente este espaço tomando-se em conta sómente o número de registros de população do censo ou pesquisa. Para um censo com um número médio de variáveis (30 a 40), se poderia armazenar cêrca de 600.000 registros de pessoas em 20 Megabytes de disco duro, ou aproximadamente 1.000.000 registros em 30 Mb.

5.3. Como parâmetro geral, se pode dizer que os discos duros são capazes de armazenar uma base REDATAM com até 3.000.000 de pessoas, o que equivaleria a 100 Mb de espaço em disco. Os arquivos que ultrapassem este limite poderiam ser dimensionados para discos óticos ("laser disk").

5.4. Em têrmos gerais, o procedimento automático de geração de bases de dados pode ser usado para um arquivo de até 300.000 registros. Os arquivos superiores a este limite, necessitam fazer uso do procedimento manual, um pouco mais complexo.

ref:\asat\red\texto\brasil.sem