SÍLVIA GAGLIARDI ROCHA

A representação documentária de informações estatísticas:
quando a ordem dos fatores altera o produto

Dissertação de Mestrado

São Paulo
2006

Para o Tião
rocha que sustenta o farol que me norteia.

Para o Lucas e a Ana Júlia
luzes que dão vida à minha existência.

Para a minha mãe
sabedoria que me é referência.

Para o meu pai
eterna fonte de inspiração.

Agradecimentos

Muitos são os que merecem meus agradecimentos pelo apoio nessa jornada, mas alguns, não posso deixar de nomear, sob pena de cometer ingrata injustiça.

Destaque especial, nunca suficientemente explicitado, para a prof^a. dr^a. Marilda Lopes Ginez de Lara, mestra, orientadora, incentivadora e, sobretudo, amiga.

Um obrigado especial também à prof^a. dr^a. Maria de Fátima Gonçalves Moreira Tálamo, pelo incentivo caloroso no momento da qualificação.

Ao Vivaldo Luiz Conti, pelo muito que aprendi nesses muitos anos trabalhando juntos e, representando a Fundação Seade, por me oferecer a oportunidade de desenvolver essa pesquisa.

A Magali e Juliette, pelas muitas leituras, conversas e paciência.

A Joice e Lucinda, amigas e companheiras em todos os passos da construção desse trabalho, sem as quais não teria chegado ao fim.

E a todos os outros, colegas de trabalho, de estudos, mestres, amigos e familiares, meu muito obrigado.

Banca Examinadora:

Relação de Quadros

Quadro 1 – Conceitos relacionados à indexação.....	p. 27
Quadro 2 – Etapas do processo de indexação.....	p. 28
Quadro 3 – Seleção de conceitos na indexação.....	p. 29
Quadro 4 – Qualidade do processo de indexação.....	p. 29
Quadro 5 – Tipos de disponibilização de informação estatística.....	p. 39
Quadro 6 – Relações básicas de um vocabulário controlado.....	p. 51
Quadro 7 – Especificidades de representação da informação estatística.....	p. 83

Relação de Figuras

Fig.1 – Uso do vocabulário na indexação e na consulta.....	p. 49
Fig.2 – Representação de relações poli-hierárquicas.....	p. 71

Resumo

Estudo sobre a representação documentária da informação estatística. A partir do reconhecimento dessa informação como representação de realidades sociais, identificam-se as suas especificidades e as da sua representação documentária visando sua divulgação na Internet e, portanto, a transmissão de informação sem a intermediação presencial do documentalista. Parte-se das premissas que a informação estatística tem características próprias que a diferenciam da informação textual, e que nem todas as recomendações para a construção de linguagem documentária, para representar esse tipo de informação, se aplicam necessariamente à representação de informações estatísticas. Utilizando referenciais da construção da linguagem documentária e da terminologia, sistematizam-se alguns parâmetros para que também o tratamento e a divulgação da informação estatística se realizem como possibilidades para a geração de novos conhecimentos.

Palavras-chave: linguagem documentária; indexação; representação documentária; disseminação da informação; informação estatística.

Abstract

This paper deals with the study on the documentary representation of statistical information. From the recognition that the statistical information represents social realities, the study identifies their specificities and their documentary representation, focusing its popularization in Internet and, therefore, its transmission without the presence of the librarian. It is based on two premises: first, the statistical information has characteristics that differentiate it from the textual information, and, second, not all the recommendations for the documentary language construction to represent textual information, necessarily represent statistical information. Using reference of documentary language construction and terminology, some parameters were systematized so that also the treatment and the popularization of the statistical information take place as possibilities of generating new knowledge.

Key words: documentary language; indexation; documentary representation; dissemination of information; statistical information.

Sumário

Resumo

Abstract

Relação de Quadros

Relação de figuras

Introdução.....	p. 01
1 – A Informação Estatística.....	p. 07
1.1 – Breve história.....	p. 07
1.2 – A história da informação estatística no Brasil.....	p. 12
1.3 - A produção da informação estatística.....	p. 18
2 – Fundamentos para a disseminação da informação estatística na Internet.....	p. 22
2.1 – Recuperação de informações.....	p. 22
2.2 – O processo de indexação.....	p. 27
2.3 – Recuperação e Linguagem Documentária.....	p. 31
2.4 – Recuperação e representação da informação estatística.....	p. 34
3 – A construção de Linguagem Documentária – aspectos teóricos.....	p. 45
3.1 – Referencial metodológico de construção de Linguagem Documentária..	p. 47
3.2 – Referencial metodológico da Terminologia.....	p. 54
4 – Linguagem Documentária para o <i>site</i> da Fundação Seade.....	p. 60
4.1 – A construção da Linguagem Documentária para dados estatísticos.....	p. 63
4.2 – Uso da Linguagem Documentária no <i>site</i> da Fundação Seade.....	p. 72
5 – Considerações finais.....	p. 81
Referências.....	p. 85
Anexos.....	p. 89

1 - Introdução

Quando pensamos em circulação ou transferência de informações, temos de considerar o cenário tecnológico de hoje, fator preponderante das modificações dos sistemas e redes de informação e comunicação, com destaque para a Internet que cria não só um fluxo contínuo de informações, mas propõe modelos novos de relacionamento com o usuário da informação documentária.

Antes do surgimento do “usuário remoto”, a interação presencial entre o usuário e o documentalista no momento da pesquisa supria eventuais falhas na comunicação entre o sistema de informação e o usuário. À medida que o documentalista orientava a pesquisa com pleno conhecimento do significado dos termos utilizados na representação e/ou baseado no seu conhecimento sobre o conteúdo das publicações, a falta de precisão na fabricação de informação documentária, ou as falhas na rede de relações paradigmáticas construídas na linguagem documentária passavam despercebidas, tanto para o usuário, como para o documentalista, já que este último, usando os recursos de memória e de conhecimento anterior, localizava a informação.

Com a Internet, a situação de comunicação se modificou, uma vez que desaparece a figura do mediador humano. Esse fato nos torna explícitos os

problemas das representações documentárias feitas segundo critérios frágeis, como coloca outras questões derivadas das formas de interação relacionadas ao uso das novas tecnologias.

No que tange à disponibilização e ao acesso à informações estatísticas as conseqüências não são diferentes. Considerando esse contexto, nos propomos, neste trabalho, analisar especificamente a informação estatística disponibilizada via Internet, considerando, portanto, a transmissão de informação sem a intermediação do documentalista, e a necessidade de uma Linguagem Documentária que permita viabilizar a transferência, dado o universo de informações por si só bastante complexo.

A Linguagem Documentária é reconhecida como instrumento construído para auxiliar na organização da informação, na sua circulação e transferência.

“Como linguagem, a LD compartilha das suas propriedades estruturais: compõe-se de unidades relacionadas, onde cada uma adquire uma significação face às demais; como linguagem construída, a LD se define como metalinguagem, isto é, uma linguagem que supõe a existência do conhecimento registrado – de uma linguagem anterior – reelaborando-o como informação. A LD não se define em relação ao acervo. Não se concebe uma LD para tratar conjuntos de registros mas sim para organizar conhecimento” (Tálamo, 1997, p.10).

A informação transmitida é, no fundo, um conjunto de conhecimentos registrados. Esse conhecimento é resultado de uma soma de *expertises* e de conhecimentos individuais que, no seu conjunto, representa a instituição que a produz. É importante salientar essa característica da informação, que é

institucionalizada, e é a instituição que lhe dá a chancela de confiabilidade e qualidade.

Ao trabalharmos a informação visando seu armazenamento e recuperação, devemos considerar que a informação tem a marca da instituição que a produz e deve ser transmitida conservando tais características. Depois de colocada em circulação, ela pode sofrer transformações de acordo com o seu uso, gerando novos estoques de conhecimento. Tais transformações é que caracterizam o movimento contínuo de geração de conhecimento.

A hipótese que defendemos aqui é a de que o tratamento dado ao um estoque de conhecimentos também carrega uma marca institucional. Quando uma instituição produz uma informação, ela tem um objetivo e um usuário específico. O trabalho de tratamento deve deixar claro, tanto quanto possível, o ponto de vista institucional sobre a informação, utilizando, para isso, instrumentos documentários específicos. Considerando, porém, que as instituições não objetivam apenas a simples reprodução de suas informações, é necessário que os instrumentos documentários permitam distintos usos e usuários da informação, segundo diferentes condições e objetivos.

Observando o contexto da circulação da informação na Internet, é preciso considerar que:

“Dois critérios permeiam o fluxo da informação entre os estoques ou espaços de informação e os usuários: o critério da tecnologia da informação, que almeja possibilitar o maior e o melhor acesso à informação disponível, e o critério da Ciência da Informação, que intervém para qualificar este acesso em termos das competências que o receptor da informação deve ter para assimilar a informação, ou seja, para elaborar a informação para seu uso, seu desenvolvimento pessoal e dos seus espaços de convivência. Não é suficiente

que a mensagem esteja disponível, ela deve também poder ser apropriada pelo receptor” (Smit & Barreto, 2002).

De acordo com essa afirmação, defendemos que a análise documentária (compreendendo as etapas de desenvolvimento da linguagem documentária, indexação, armazenamento e recuperação da informação) também deve ser reconhecida como instância de construção de conhecimento.

Sob essa perspectiva, e considerando as características institucionais da informação, é preciso dispor de mecanismos que levem à apropriação, mecanismos esses que não se confundem com as linguagens universais de tratamento da informação. Ao contrário, uma linguagem documentária que funcione como instrumento de conhecimento tem características próprias que combinam o modo institucional de conceber a informação, mais as possibilidades de sua reutilização.

O universo de informações estatísticas tem características próprias que a diferenciam da informação textual. Neste trabalho, portanto, procuramos identificar elementos para compreender melhor o que são os dados estatísticos e pesquisar as possibilidades de sua representação em ambientes digitais. Utilizando referenciais de construção da linguagem documentária e da terminologia, propomo-nos a identificar alguns parâmetros para que o tratamento e a divulgação da informação estatística se realize como oportunidade para a geração de novos conhecimentos.

No primeiro capítulo procuramos fazer uma breve história da informação estatística e, especificamente, da estatística brasileira, ressaltando aspectos atuais da produção das informações estatísticas.

No segundo capítulo discutimos a disseminação da informação estatística na Internet. Para efeito de contextualização, iniciamos o capítulo com uma discussão sobre recuperação de informação, a partir das reflexões de Ingwersen, que foca a recuperação da informação em sistemas eletrônicos. Apesar desses estudos terem sido desenvolvidos anteriormente à explosão da Internet, os conceitos mapeados pelo autor e as dificuldades por ele apontadas, relativas aos sistemas de recuperação, se mantêm como norteadores para os que se propõem a contribuir para a melhoria da circulação da informação e da produção do conhecimento.

A segunda parte do capítulo destaca os principais conceitos do processo de indexação.

A terceira parte faz uma breve explanação sobre a linguagem documentária e as possibilidades que ela permite na resolução dos problemas de recuperação de informação levantados no início do capítulo.

A quarta parte fecha o capítulo relatando questões pertinentes ao atual estágio em que se encontram as questões de recuperação e representação de informação estatística, sugerindo-se o uso de linguagem documentária para auxiliar na resolução dos problemas levantados.

O terceiro capítulo, dividido em duas partes, abrange os aspectos teóricos e metodológicos da construção da linguagem documentária e do uso dos referenciais da Terminologia, que, combinados, constituem a base para o estabelecimento de procedimentos de organização de informação visando sua recuperação.

No capítulo quarto, descrevemos o processo de construção da linguagem documentária utilizada na indexação de informações estatísticas disponibilizadas no *site* da Fundação Seade, instituição responsável pela produção e divulgação das estatísticas do governo do Estado de São Paulo. No bojo dessa discussão abordamos também questões importantes sobre as especificidades da informação estatística que interferem diretamente na construção da linguagem.

No processo de desenvolvimento dessa linguagem e de sua aplicação na indexação surgiram novas questões: algumas relacionadas ao processo de indexação particular das informações estatísticas; outras, relacionadas à especificidade de seu sistema de recuperação.

Nas considerações finais, procuramos sistematizar a experiência, de forma a possibilitar a indicação de alguns parâmetros não só para a construção de linguagem documentária, como também para seu uso no tratamento e na recuperação da informação estatística.

1 - A informação estatística

1.1– Breve história

A informação estatística tem sua origem ligada ao Estado e a necessidade de inventariar bens, produções e pessoas, com fins arrecadatários e militares, pois era preciso saber quantos soldados estavam disponíveis para os nem tão eventuais conflitos, havendo portanto, mil justificativas para se contar e controlar essa população.

Naturalmente, essas contagens eram rudimentares e não necessariamente exaustivas, não podendo ser consideradas como recenseamento, que é definido como:

“Estudo científico de um universo de pessoas, instituições ou objetos físicos com o objetivo de obter conhecimentos quantitativos acerca das características importantes dessa população” (Ferreira & Tavares, p. 4).

Segundo Matos, em seu *site* sobre estatística, e também segundo página sobre a História da Estatística no site da Universidade do Rio Grande do Norte, a palavra **censo** tem sua origem na palavra *CENSERE*, que em latim significa **taxar**.

No IBGE, no *site* sobre o Censo 2000, na História do Censo no Mundo, encontramos o seguinte:

“A palavra Censo se origina do latim *census* que quer dizer conjunto dos dados estatísticos dos habitantes de uma cidade, província, estado, nação etc”.

Mas, independente da interpretação dada à palavra *censo*, a verdade é que os primórdios da informação estatística perdem-se na história antiga.

Deparamo-nos com alusões a recenseamentos efetuados no Egito, datados de 2900 a.C., assim como se reconhece o princípio da Estatística Descritiva¹ na obra de Aristóteles, que viveu entre o ano 384 e 322 a.C. Existem também registros de recenseamento no Antigo Testamento, no livro do Êxodo (30, 12-15), na Índia, no Japão e na China, onde registros históricos dizem que o primeiro recenseamento ocorreu no ano 2238 a.C. (Ferreira & Tavares, p. 8).

Apesar de os primeiros recenseamentos serem tão antigos, manteve-se a mesma prática de coletar dados sobre produção e população, com fins militares e de taxaço, até relativamente pouco tempo.

Durante a idade média nasceu a Teoria das Probabilidades, resultante das tentativas de avaliar possibilidades em jogos de azar. Girolano Cardano (1501-1576) publicou um pequeno manual de jogos de azar “*Líber de Ludo Aleae*”, introduzindo as técnicas combinatórias, mas nunca chegou a produzir um teorema, ficando o reconhecimento da instituição do Cálculo das Probabilidades, com Pascal (1632-1662) e Fermat (1601-1665) (Ferreira & Tavares, p. 14).

Somente nos séculos XVII e XVIII a estatística começou a se aproximar da ciência que conhecemos hoje, principalmente com o aparecimento de duas Escolas, uma na Alemanha e outra na Inglaterra.

¹ Estudo descritivo de dados de uma amostra (ou de uma população) em que se resume toda a informação recolhida em gráficos e tabelas, calculando algumas das suas características, por exemplo, a moda, a média, etc. (Ferreira & Tavares, p.5)

A Escola alemã ficou conhecida como Escola Descritiva, mas se afastou das idéias que fundamentam a estatística moderna; entretanto, ao seu representante mais famoso, Gottfried Achenwall (1719-1772), se atribui a cunhagem da expressão “estatística”.

Existe porém, sobre o uso dessa expressão, uma outra versão apontada por Ferreira & Tavares, que diz:

“... na opinião de Sir Maurice Kendall (Pearson e Kendall, 1820), esta palavra já tinha sido utilizada em Itália, num trabalho do historiador Girolamo Ghilini, em 1589 que se refere a um registo da “*civile, politica, statistica e militare scienza*” (...) Segundo Kendall, a palavra utilizada na Escola Alemã denotava apenas o método utilizado nos estudos dedicados à descrição dos estados políticos e, se alguma informação numérica aparecia era somente por acaso ou conveniência” (Ferreira & Tavares, pp.16-17).

Senra (1999), quando atribui a criação da expressão à escola alemã, ressalta que na época estatística era considerada como:

“*ciência do Estado* ou como a ciência *que se refere ao Estado*. Mais precisamente, referia-se aos acontecimentos tidos como memoráveis ao entendimento de um Estado, descrevendo-se seu território e sua população, compondo assim referências a amparar a ação de seus dirigentes”

“... semelhante aos trabalhos de geógrafos e historiadores, nesses documentos, os números não são predominantes, seja por não estarem sistematicamente disponíveis, seja também por não serem considerados essenciais a uma boa explanação...” (Senra, 1999).

A Escola inglesa, “Escola de Aritméticos Políticos”, preocupava-se com o estudo numérico dos fenômenos sociais e políticos e, do trabalho desenvolvido por seu representante, John Graunt (1620-1674), constituiu-se a base da Estatística Moderna.

Graunt foi o primeiro a estabelecer métodos de pesquisa estatística:

“estudou a mortalidade em Londres e as incidências das causas naturais, sociais e políticas desse fenômeno”... “foi a primeira pessoa a fazer observações entre sexos e mostrou que nasciam mais homens que mulheres e que por cada 100 pessoas nascidas, 36 morriam aos 6 anos e 7 sobreviviam até os 70 anos” (Ferreira & Tavares, p.17).

A partir do trabalho de Graunt, que deu “grande impulso à análise quantitativa dos fenômenos sociais e ao desenvolvimento das Estatísticas Demográficas” (Ferreira & Tavares), muitas pessoas e estudos contribuíram para o desenvolvimento da Estatística, transformando-a, de simples arrolamento de dados, na atual Ciência Estatística, utilizada na formação de políticas públicas, na quantificação e qualificação das condições de vida da população, na medida de outras dimensões da realidade social, e na possibilidade de controle da sociedade sobre o estabelecimento e aplicação das políticas públicas, estando finalmente também a serviço do cidadão e não somente do Estado.

Nos dias atuais, considera-se que:

“... as estatísticas são representações numéricas da realidade a qual buscam mensurar. Em seu processo de construção, apóiam-se em interpretações teóricas que modelam aspectos da realidade e passam a criar seus próprios modelos de interpretação do real. As estatísticas que buscam apreender a realidade social guardam, pois, em sua configuração numérica, uma certa visão do mundo, o que as restringe ao contexto que as referencia” (Porcaro, 2001, p.2).

“As estatísticas não refletem a realidade, refletem o olhar da sociedade sobre si mesma” (Benson, 1995, p.19).

Vemos, portanto, que as estatísticas representam uma leitura da configuração social, leitura essa criada para representar uma realidade. Sob esse ponto de vista, dado estatístico é uma tradução.

“... resulta de um trabalho de conceituação, de organização, de observação, de exploração, e essas operações têm um custo. Não se produz, então, informação sem uma demanda correspondente, expressa nas formas requisitadas, mercantis para os produtores privados, institucionais para o produtor público” (Benson, 1995, p.41).

1.2 – A história da informação estatística no Brasil²

Na história da estatística brasileira, mesmo reconhecendo que antigos relatórios não podem ser considerados como testemunhas da existência de estatísticas, organizadas, elaboradas e com continuidade, poderíamos considerar como primeiro arrolamento das coisas do Brasil, o relato de Vaz de Caminha a D. Manuel sobre os variados aspectos observados na terra de Santa Cruz.

Prosseguindo por essa linha iríamos encontrar dados oriundos das expedições de exploração do litoral, contagens de Anchieta a quantificar os componentes do primeiro rebanho colonizador, os índios civilizados e os escravos africanos, ao todo, 57.000 almas, e cálculos de Nóbrega sobre o número de povoadores que era mister multiplicar.

Encontraríamos informações gerais das capitanias e os estudos de missionários, naturalistas e bandeirantes; chegaríamos, depois, ao cômputo da população do Rio de Janeiro em 1799 e ao censo geral de 1808, esse, ordenado por D. Rodrigo de Souza Coutinho, com fins puramente militares.

Poderíamos relatar a lei de 1826 preconizando tabelas estatísticas uniformes para todas as Províncias e nos deteríamos nas iniciativas de Luiz Maria da Silva Pinto, em Minas Gerais, nas de Daniel Pedro Müller, em São Paulo, que inclusive produziu um ensaio sobre o assunto.

² Para evitar inúmeras citações repetidas, registro que a maior parte desse tópico foi obtida a partir de resumos de capítulos da obra “A administração Pública e a Estatística”, de Germano G. Jardim, 1941.

Teríamos de citar o decreto que criou a Comissão de Estatística Geográfica Natural, Política e Civil em 1829, e o que a extinguiu em 1834, bem como regulamentos e outros atos dispendo sobre a organização das estatísticas no Império.

Verificaríamos outras contribuições, como a de Francisco de Paula Cândido, que organizou as primeiras estatísticas de letalidade no país, assim como, as “Bases para a organização da estatística”, apresentadas em 1862 ao ministro da Agricultura por José Cândido Gomes, e também a atuação de Sebastião Ferreira Soares, autor de “Elementos da Estatística”, em 1865.

Chegaríamos então em 1870, ano da criação da Diretoria Geral de Estatística, que coordenou as primeiras atividades sistematizadas, ligadas a levantamentos censitários, culminando em 1872, com o primeiro Recenseamento Geral do Império do Brasil.

Destacamos aqui, mais como fato curioso, do que parte da cronologia da história da estatística brasileira, o uso feito por Rui Barbosa das estatísticas escolares, que em pleno florescimento em muitas partes do mundo, permitia comparações mais expressivas.

Rui Barbosa, em o “Parecer de 1822”, cruamente denunciava à Câmara Legislativa

“a ignorância nacional e a inconsciência geral do estado de profundas trevas que afogam o espírito do povo brasileiro” (Barbosa, 1883, *apud* Jardim, 1941, p.25).

Foram os números a sua mais poderosa arma para demonstrar os achaques da nossa formação social naquela época e justificar a reforma do ensino que a comissão de que era relator propunha ao governo. O uso dos números como argumentação não o impediu de enunciar:

“Antes de mais nada, é óbvio que a estatística escolar, nas verdadeiras condições de segurança e clareza impostas a esse serviço pelas exigências da verdade, está por criar nesse país, onde até hoje, os resultados obtidos nesse ramo da administração se ressentem de uma grosseria, de uma obscuridade, de uma confusão, de uma incongruência difíceis de fazer sentir a quem não os tenha examinado com a paciência minuciosa com que os esquadrinhamos” (Barbosa, 1883, *apud* Jardim, 1941, p.25).

Essas palavras demonstravam quão pouco profícua fora a ação coordenadora da primeira repartição oficial de estatística, o que a levaria à extinção em 1879.

Após a mudança de regime, em novembro de 1889, as necessidades de conhecimento da vida nacional produziram, como um dos primeiros atos da recém nascida República, o restabelecimento da Diretoria Geral de Estatística, por decreto de 2 de janeiro de 1890. A Diretoria tinha a incumbência de elaborar estatísticas sobre a população, indústria e comércio, justiça, finanças, beneficências, transportes e comunicações, administração e serviços públicos, instrução, organizações sócio-religiosas, ordem e defesa pública, além de correspondência com organizações congêneres do estrangeiro.

Também em 1890 temos o segundo recenseamento geral do país, o primeiro da República.

Em 1891 aparece o primeiro anuário de estatística demógrafo - sanitária.

O século XX inicia-se com a reorganização dos serviços de estatística comercial, que ficam a cargo de um órgão do Ministério da Fazenda. As estatísticas de demografia sanitária desenvolvem-se de modo animador, mas os trabalhos do Censo de 1900 não correm satisfatoriamente, e seus resultados só são conhecidos em 1907, ano em que se inicia a remodelação da Diretoria Geral de Estatística, sob a direção de Bulhões de Carvalho.

Em 1910 é suspenso o recenseamento geral que se realizava naquele ano, em função da Revolta da Chibata (Correio Braziliense, 2001).³

Depois disso, o sistema estatístico federal ainda sofre sucessivas modificações, mas as atividades estatísticas da União continuam a se desenvolver através de inquéritos de vulto e em alguns Estados já é grande a preocupação de aparelhar melhor as administrações locais com os serviços estatísticos.

A guerra de 1914 faz sentir seus efeitos na vida econômica deste hemisfério, e, em nosso país, a necessidade de coordenação mais eficaz das informações estatísticas leva, em 1915, a nova reorganização do departamento de Estatística.

Em 1916 é publicado o primeiro número do Anuário Estatístico do Brasil (volume I), referente ao período de 1908 a 1912, contendo dados sobre o Território e População. Um ano depois é editado o volume II relativo a Economia e Finanças, e em 1927, conclui-se a obra com o volume III, trazendo informações sobre Cultos, Assistência, Repressão e Instrução.

³ Informação obtida no site http://www2.correioweb.com.br/cw/2001-12-20/mat_25431.htm - acessada em março de 2004.

Em 1920 foi realizado o quarto censo, abrangendo não só múltiplos aspectos das condições econômicas do país, como também a mais exaustiva descrição demográfica até então realizada na América Latina.

Durante a preparação do censo de 1930, que não se realizou por causa do delicado momento político que a nação vivia, gravitava, no pensamento dos nossos realizadores da estatística, uma ansiedade por uma renovação do processo de política pública, que erigisse, como princípio básico a cooperação das três ordens administrativas em um sistema genuinamente brasileiro.

Estavam lançadas as sementes da racionalização da estatística nacional, e a grande campanha encetada por Bulhões de Carvalho e continuada por Teixeira de Freitas só necessitou aguardar o momento propício para frutificar.

Em 1934 foi criado o Instituto Nacional de Estatística, que só existiu de fato em 1936, mudando em 1938 para Instituto Brasileiro de Geografia e Estatística – IBGE.

“Com a criação do IBGE e com a contribuição do renomado demógrafo italiano, Giorgio Mortara, inaugurou-se a partir do censo de 1940, a moderna fase censitária no Brasil, caracterizada pela periodicidade decenal. As pesquisas dos censos tiveram sua abrangência temática ampliada com a introdução de quesitos de interesse econômico e social, tais como os de mão de obra, emprego, desemprego, rendimento, fecundidade, migrações internas, etc.” (IBGE, 2000).

A história do IBGE, assim como a da produção estatística, continuou apresentando altos e baixos, flutuando ao sabor das crises econômicas e políticas que permearam o país.

Como apontam Guizzardi & Conti (2001), na década de 80 a deterioração das finanças públicas, a explosão da inflação, o enfraquecimento e desorganização das ações de governo refletiram duramente sobre a produção das estatísticas, seja pela não utilização desse instrumento para o planejamento, seja pela diminuição dos recursos a ela destinados. Com o passar dos anos, cortes orçamentários, aviltamento de salários provocando perda de técnicos qualificados, a não reposição de quadros funcionais e outras situações similares, foram minando a produção das estatísticas, chegando a ponto da não realização do Censo Demográfico de 1990, quebrando a regularidade mantida desde 1940, e que só a duras penas foi realizado em 1991.

A reconstrução do IBGE foi iniciada nos anos seguintes com um movimento no sentido de busca da modernização das pesquisas realizadas incorporando-se modernas técnicas de amostragem, buscando-se a cooperação de organismos internacionais e paralelamente, providências como reposição de quadros, recuperação da infra-estrutura e aquisição de novos equipamentos, vão acelerando a recuperação do IBGE e o coloca, já há alguns anos, em um novo ciclo virtuoso.

1.3 - A produção da informação estatística

A produção das estatísticas no Brasil continua fortemente apoiada na produção do IBGE, cuja história em traços muito rápidos vimos no tópico anterior.

A principal fonte de dados estatísticos é o Censo Demográfico, realizado em intervalos decenais em todo o território nacional. Isso se deve não só à regularidade dos Censos (exceção feita ao levantamento de 1990, efetuado em 1991), como também à ampliação do leque temático da pesquisa. Outra característica importante do Censo é sua ampla cobertura geográfica.

O fato de os resultados dos censos poderem ser recuperados em diversos níveis de desagregação, indo da totalidade do território até ao setor censitário – unidade de coleta que, na zona urbana, compreende cerca de 300 domicílios – o transformam não só na principal, mas na quase exclusiva fonte de informações para o planejamento de políticas em âmbito local e micro-local (distritos, bairros, etc.), uma vez que as outras fontes não gozam de confiabilidade ou não têm cobertura espacial ou populacional muito abrangente (Jannuzzi, 2001).

Além das pesquisas censitárias, as estatísticas vêm sendo produzidas não só pelo IBGE, mas também pelas agências estaduais de estatísticas, através de pesquisas amostrais, concebidas para prover o país de informações demográficas e socioeconômicas, nos intervalos censitários, ou de pesquisas com temática específica como as que medem o mercado de trabalho e uma série de outras investigações realizadas, com o intuito de aprofundar a abrangência temática pesquisada e complementar cronologicamente o levantamento censitário.

As estatísticas coletadas por meio de censos demográficos, pesquisas amostrais ou registros administrativos

“... constituem-se na matéria prima para a construção de indicadores sociais. O que diferencia a estatística pública do indicador social é o conteúdo informacional presente, isto é, ‘o valor contextual’ da informação disponível neste último” (Jannuzzi, 2001, p.16).

Ao analisar a estatística produzida no país, desde o começo do século XX, percebe-se claramente uma alteração gradativa quanto ao tipo de informação pesquisada. A produção das estatísticas, nos dias de hoje, é muito diversificada nos quesitos que são investigados e muito detalhada na abrangência geográfica solicitada.

Do mesmo modo, nos últimos anos a demanda por informações conhece um acentuado crescimento, que direta ou indiretamente afeta as instituições produtoras de informação estatística.

Verifica-se que:

“Atualmente, as entidades produtoras de informação vêem-se alvo das mais distintas e crescentes demandas por informações. Isso as tem obrigado a se reorganizar com vista em atendê-las de modo adequado e permanente. São vários os elementos que explicam esse fenômeno...”

“... Diante das profundas mudanças por que nosso país tem passado, seria impossível, nesse espaço, tratar de todas elas e de seus impactos sobre a produção de estatísticas. Assim, destacaram-se algumas, talvez as mais visíveis, mas que, certamente, revelam apenas parte do problema” (Ferreira, 2003, p.18).

Em resumo, a análise feita por Ferreira (2003) destaca as seguintes razões que afetam a produção das estatísticas no Brasil:

- a) O processo de democratização: governos democráticos procuram desenhar políticas com base em informações estatísticas, pois prestam conta de suas ações à sociedade que os elegeu e ampliam os órgãos de controle e avaliação dessas ações - entram em cena novos usuários de informação estatística, pois partidos políticos, organizações não governamentais, sindicatos e outras organizações sociais necessitam de informações para definir suas atuações ou acompanhar as ações governamentais.
- b) Descentralização da ação pública: processo que se acentuou a partir dos anos 80, como uma espécie de contraponto à grande centralização do poder político e da ação governamental do período autoritário, provocou, principalmente na década de 1990, um movimento de municipalização, do qual se espera que a proximidade entre os governantes e os cidadãos signifique uma maior presença da sociedade na definição, no monitoramento e na avaliação das políticas públicas - é necessário que os cidadãos disponham de instrumentos para que isso se concretize, surgindo a necessidade de um grande conjunto de informações municipais e, em muitos casos, intramunicipais.
- c) Focalização das políticas sociais: várias ações governamentais passaram a eleger segmentos específicos da população como público alvo, requerendo para tanto informações detalhadas das características da população, de modo a permitir a identificação do segmento prioritário, objeto da intervenção, incluindo, em muitos casos, sua localização espacial e a construção de cadastros das famílias ou pessoas beneficiadas.

“Um aspecto adicional, porém, merece ser destacado. Para que subsidiem a elaboração, acompanhamento e a avaliação das ações governamentais, as informações estatísticas, além de cobrir amplo escopo temático e territorial, necessitam ser atualizadas permanentemente. Isso significa que não bastam informações censitárias, atualizadas a cada dez anos, nem as originárias de pesquisas amostrais, que são mais freqüentes, mas com possibilidades de desagregação espacial limitadas. Assim, as instituições produtoras de informações têm de valer-se ou de levantamentos primários específicos às necessidades de seus usuários – cujo custo, em geral, é muito elevado – ou da utilização de registros administrativos que, originalmente, foram construídos para outros fins. Expandem-se, assim, suas tarefas de permanente coleta, organização, avaliação, validação e disponibilização desses registros, ampliando significativamente o escopo de atuação dessas instituições” (Ferreira, 2003, p.19).

Outro grande desafio enfrentado pelas instituições produtoras de estatísticas está relacionado com a disseminação dessa informação.

A partir do momento que essa informação passa a ser demandada por variados segmentos da sociedade, e não só por especialistas, e que os progressos das tecnologias de informação e comunicação começam a ser utilizados por um público muito mais amplo, as instituições produtoras de estatísticas vêem-se frente à tarefa de conseguir disponibilizar suas informações de forma rápida, com baixo custo e com facilidade de acesso. A esse enfrentamento, a Internet se destaca como grande ferramenta para solucionar questões de custo, velocidade de transmissão de grandes volumes de dados e facilidade de acesso. No entanto, o encaminhamento das questões relativas à efetiva transmissão de informação ao usuário leigo ainda continuam muito aquém do desejado, como discutiremos no decorrer deste trabalho.

2 – Fundamentos para a disseminação da informação estatística na Internet

2.1 - Recuperação de Informações

A recuperação de informações - RI é um dos campos de pesquisa centrais na Ciência da Informação (Jarvelin e Vakkari, 1992, *apud* Ingwersen, 1992). Preocupa-se com os processos envolvidos na representação, armazenamento, busca e recuperação de informação relevante para um pedido de informação feito por um usuário humano (Ingwersen, 1992).

As reflexões de Ingwersen referem-se à recuperação da informação em um contexto ditado pela tecnologia, ou seja, focam a recuperação de informação em sistemas eletrônicos. Necessário dizer, entretanto, que a preocupação com a recuperação de informação é bastante antiga, sendo emblemática a contribuição de Otlet e La Fontaine que, no final do séc XIX, já defendiam a circulação e o acesso social à informação.

“Otlet conferiu à documentação um papel social da maior relevância, uma vez que o mesmo priorizava em suas análises a informação como processo, a informação gerando o conhecimento, a informação diminuindo a insegurança e a incompreensão entre os homens” (Tálamo et al., 2002).

Independentemente do fato de verificar que a recuperação da informação já era importante antes das facilidades tecnológicas, hoje, quando pensamos em circulação ou transferência de informações, temos de considerar o cenário tecnológico atual, fator preponderante das modificações dos sistemas e redes de informação e comunicação, com destaque para a Internet, que cria não só um fluxo

contínuo de informações, mas propõe modelos novos de relacionamento com o usuário de informação.

Ingwersen (1992), em seus estudos sobre recuperação de informações, destaca seu entendimento de informação baseado em estudos de Salton, Shannon, Artandi, Wersig, Belkin e M. De Mey, propondo que o conceito, na perspectiva da Ciência da Informação, deva ser visto como indo além do significado: informação é uma transformação de estruturas de conhecimento dos geradores em forma de signos (que estruturados podemos traduzir por linguagem), e quando apreendida, um transformador do estado de conhecimento do usuário.

Esta conceituação vem ao encontro não só do pensamento já citado de Otlet, mas é hoje razoavelmente consensual em estudos de Ciência da Informação. Simplificando, podemos entender informação como sendo modificadora do estado de conhecimento, e a recuperação da informação como um processo que permite essa modificação.

Ao lado dessa visão podemos associar outra, não menos importante, ou seja, a proposta de ver o conhecimento em si mesmo como estoque, enquanto informação é conhecimento comunicado, é fluxo (Barreto, 1994). Essa associação reforça o reconhecimento dos sistemas de recuperação de informação como essenciais para a Ciência da Informação, pois são eles que fazem a mediação entre geradores e receptores de informação, contribuindo ao processo de geração de conhecimento.

Os estudos em recuperação da informação, mapeados por Ingwersen (1992), identificaram claramente que o grande problema da recuperação da informação está centrado na significação, salientando que alguns textos são mais

relevantes a um pedido específico por informação do que outros, e que um texto específico pode ter uma significação diferente para vários pedidos de informação.

Para o autor, três áreas básicas, inter-relacionadas, constituem o âmago dos problemas de recuperação de informações: *aboutness*; representação e relevância.

Aboutness se refere a “sobre o que é o documento”, podendo-se encontrar o conceito sob as seguintes variações: *aboutness* do autor; *aboutness* do indexador; *aboutness* do usuário; *aboutness* do pedido. Para Ingwersen, as questões de representação estão diretamente ligadas ao tipo de *aboutness* e a relevância é um valor de natureza pragmática, ligado a cada usuário e a seu estado de conhecimento.

J. van Rijsbergen (apud Ingwersen, 1992) fala das relações entre *aboutness* e relevância em recuperação de informação, considerando a relevância como “a medida ou grau de correspondência ou utilidade existente entre um texto ou documento e uma pergunta ou busca de informação como determinado por uma pessoa”.⁴ O autor destaca como característica fundamental da recuperação da informação o fato de que sua eficiência está longe de atingir 100%. A pessoa que busca, o próprio sistema de recuperação, assim como o pesquisador que estuda seus fenômenos, cada qual “não sabe o que ele não recupera” e nunca saberá. Essa absoluta incerteza de recuperação, inerente do ponto de vista cognitivo, pode se constituir no único princípio universal original na Ciência da Informação.

⁴ Entende-se a pergunta como uma pesquisa específica quando o usuário busca a informação com características já conhecidas e pré-determinadas por ele; já a busca de informação é entendida como sendo uma pesquisa genérica, realizada quando o usuário quer saber quais informações existem sobre determinado assunto.

A partir dessas afirmações sobre as questões de significação e representação, podemos propor uma aproximação entre *aboutness* e linguagens, para verificar a partir de que referências se afirma “sobre o que é o documento”:

- *Aboutness* do autor = Linguagem do especialista;
- *Aboutness* do indexador = Linguagem documentária;
- *Aboutness* do usuário = Linguagem natural;
- *Aboutness* da pergunta = palavras com que os usuários fazem suas buscas.

Considerando esse paralelo, os tipos de *aboutness* podem ser relacionados a diferentes perspectivas sobre o mesmo documento:

- *Aboutness* do autor – perspectiva de quem vê a informação – o autor ou a instituição produtora da informação;
- *Aboutness* do indexador – perspectiva da direção para onde se orienta a informação – tratamento do dado;
- *Aboutness* do usuário – perspectiva da relevância – conceito que varia de acordo com o usuário.⁵

Quanto ao que diz respeito aos problemas de recuperação e relevância da informação levantados por Ingwersen, sugerimos que uma solução viável para minimizar essas questões está no uso de linguagens documentárias que têm como função organizar conhecimento de modo a possibilitar sua transformação em informação no processo de sua transferência, com o intuito de gerar novos conhecimentos.

⁵ Acreditamos que o *aboutness* da pergunta está pressuposto no *aboutness* do usuário. Há uma sutil distinção entre eles: o primeiro se refere a pergunta em linguagem natural e o segundo, à busca através de palavras isoladas, que não chegam a ser um enunciado ou a explicitar um conceito, mas que se referem a uma idéia que corresponde a um assunto genérico.

A linguagem documentária, citada no parágrafo anterior como encaminhamento da solução para as indagações de Ingwersen, é o instrumento que deve ser usado para a indexação dos documentos contidos no sistema, e consubstancializa concomitantemente, o elo entre os diversos tipos de *aboutness*, pois a linguagem documentária, quando de sua elaboração, tece uma teia de relações conceituais entre seus termos e se coloca como uma linguagem intermediária. É intermediária em vários sentidos: ponte entre a linguagem do sistema e a linguagem do usuário (Lara,1998) e, mais ainda, ponte entre as diferentes linguagens que dão forma à informação. Em outras palavras, a linguagem documentária possui recursos que garantem a relação entre as linguagens dos especialistas, dos indexadores e dos usuários, que como dissemos anteriormente, estão relacionadas diretamente com os vários *aboutness*.

As questões de relevância e pertinência devem ser objeto de cuidados na etapa do processo de análise documentária que envolve a determinação de conceitos e seleção de termos de indexação que compõem o assunto.

2.2 - O Processo de Indexação

Para melhor entendimento do processo de indexação, salientamos aqui os principais conceitos referidos na NBR12676:1992, que trata dos métodos para a análise de documentos e determinação de seus assuntos e seleção de termos de indexação.

Quadro 1 – Conceitos relacionados à indexação

Indexação – Ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação.

Documento – Qualquer unidade impressa ou não, que seja passível de catalogação ou indexação.

Conceito – Qualquer unidade de pensamento. O conceito pode ter o seu conteúdo semântico reexpresso pela combinação de outros conceitos, que podem variar de uma língua ou de uma cultura para outra.

Assunto – Tema representado num documento por um conceito ou combinação de conceitos.

Termo de indexação – Representação de um conceito sob uma das seguintes formas:

- a) Termo derivado da linguagem natural, de preferência um nome ou locução nominal;
- b) Símbolo de classificação.

Fonte: NBR12676:1992

O conceito central para a nossa análise é o de indexação. A aplicação da norma de indexação à informação estatística será objeto de discussão no capítulo 4.

Ainda segundo a norma citada acima, o processo de indexação consiste basicamente de três estágios:

Quadro 2 – Etapas do processo de indexação

- exame do documento e estabelecimento do assunto de seu conteúdo
- identificação dos conceitos presentes no assunto
- tradução dos conceitos nos termos de uma linguagem de indexação

Fonte: NBR12676:1992

A primeira etapa, onde se estabelece o assunto, deve ser feita através de uma análise do documento, não se descartando a hipótese de consultar bibliografia auxiliar ou até um especialista na área em questão, para garantir uma perfeita delimitação do conteúdo.

Na segunda etapa, o indexador deve identificar os conceitos que são elementos essenciais na descrição do assunto. Assim, se o documento discorre sobre a importância dos exames pré-natais na redução da mortalidade infantil, na década de 90, em São Paulo, os conceitos envolvidos são: exames pré-natais, mortalidade infantil, década de 90 e São Paulo. Além disso, o indexador deve também identificar outras características, verificando: se a abordagem do assunto é metodológica; se os resultados relatados são fruto de pesquisa de campo; se a motivação do estudo foi, por exemplo, do ponto de vista de políticas públicas ou se foi sociológico, etc.

A norma ainda pondera:

Quadro 3 – Seleção de conceitos na indexação

Na seleção de conceitos, o principal critério deve ser sempre o valor de um conceito para a expressão e recuperação do assunto do documento. Ao fazer a escolha dos conceitos, o indexador deve ter em mente as consultas que podem ser feitas ao sistema de informação. Neste contexto, o indexador deve:

- a) Escolher os conceitos que forem considerados os mais apropriados para uma determinada comunidade de usuários;
- b) Adaptar tanto os instrumentos de indexação como os próprios procedimentos em função da retroalimentação obtida através dos pedidos de informação. Essa adaptação, entretanto, não deve alterar a estrutura lógica ou a lógica da linguagem de indexação.

Fonte: NBR12676:1992

A qualidade do processo de indexação está relacionada a fatores de:

Quadro 4 – Qualidade do processo de indexação

a) Consistência na especificidade dos termos atribuídos:

a escolha dos conceitos para a representação afeta diretamente as características de exaustividade e especificidades, e por esse motivo, deve ser relacionado a finalidade do sistema de recuperação para o qual se está indexando.

b) Qualificações do indexador (imparcialidade, conhecimento, etc):

a imparcialidade é um fator necessário para se obter consistência na indexação, que é dificultada também quando se tem uma equipe muito grande de indexadores, e também, quando os níveis de conhecimento do assunto ou dos procedimentos de indexação não são adequados. Recomenda-se, nesses casos, uma etapa de verificação, cujos resultados devem ser repassados aos indexadores.

c) Qualidade dos instrumentos de indexação:

a hospitalidade da linguagem documentária utilizada, interfere na qualidade da indexação. Esta deve admitir livremente novos termos ou mudanças na terminologia, bem com atender a novas necessidades dos usuários.

Fonte: NBR12676:1992

A referência à norma, aqui destacada, tem como objetivo não só fixar os conceitos relativos à indexação úteis ao processo de tratamento da informação estatística, como também verificar quão suficientes e adequadas eles são quando aplicados a esse gênero de informação.

Retomaremos a questão no capítulo 4.

2.3 – Recuperação e Linguagem Documentária

A linguagem documentária – LD se propõe como instrumento que realiza a compatibilização entre a linguagem do sistema e a do usuário. É o instrumento que, por excelência, trata a informação, tornando-a disponível para circulação e transferência.

“A função da LD é tratar o conhecimento dispondo-o como informação. Em outras palavras, compete às LDs transformar estoques de conhecimento em informações adequadas aos diferentes segmentos sociais. É esse partilhamento que está na base do caráter público da informação e que não pode ser obtido na ausência de uma LD” (Cintra, et al. 2002, pp. 16-17).

O partilhamento de conhecimentos entre os diversos agentes envolvidos na recuperação da informação só se concretiza no momento em que o estoque de informação é recuperado e apreendido pelo usuário, ou seja, quando a comunicação é efetuada. É, também, resultante do entrecruzamento de distintas linguagens. Como destaca Lara,

“A partir do enfrentamento das várias linguagens - a dos especialistas, a dos utilizadores de informação -, são modelizadas as linguagens documentárias, instrumentos que permitem a materialização da representação de informações, de forma a garantir a comunicação. As linguagens documentárias funcionam, assim, como meio de estabelecer as pontes que viabilizam o acesso e uso da informação” (Lara, 1998, p. 100).

A expressão linguagem documentária - entendida como uma ferramenta construída para cumprir essa função de traduzir o conteúdo dos documentos da linguagem natural para uma linguagem que não só representasse as informações contidas nos documentos, mas também as organizasse para fins de recuperação -

apareceu na década de 1970, apesar de já existirem ferramentas construídas para esse fim.

“De maneira geral, define-se linguagem documentária (LD) como uma linguagem construída, oposta à natural, portanto, que tem como objetivo específico tratar a informação para fins de recuperação” (Tálamo, 1997, p.10).

As primeiras ferramentas conhecidas, e de ampla utilização, porém não caracterizáveis exatamente como linguagem, foram os sistemas de classificação bibliográfica CDD - Dewey Decimal Classification e CDU - Classificação Decimal Universal. Esses sistemas são de natureza enciclopédica: visam cobrir toda a gama do conhecimento humano. Mais tarde, apareceram os sistemas de classificação facetada, que já visavam cobrir domínios particulares. A partir das classificações facetadas, desenvolveram-se os tesauros, que incluem uma preocupação adicional: a do controle do vocabulário. É a partir desse momento que se constitui realmente uma linguagem, pois aparece a preocupação com as significações dos termos e a identificação das suas relações, o que abre um grande leque de possibilidades de representar e de recuperar a informação com maior consistência.

Nessa perspectiva, conclui-se que a elaboração de uma linguagem documentária envolve a identificação de mecanismos para organizar conhecimento de modo a possibilitar sua transformação em informação no processo de sua transferência, objetivo último da Ciência da Informação.

A linguagem documentária, com sua rede de relações de significados, disparando processos semióticos orientados pelos seus termos genéricos, termos específicos, termos de equivalência e com o uso de notas explicativas e definições, constrói pontes entre diversas linguagens, de diversas áreas de especialidades e usuários diversos.

Sob outro aspecto, essas estruturas funcionam, de certo modo, como mapas cognitivos que orientam o uso e a recuperação de informações, o que reforça a hipótese de que a abordagem da Análise Documentária (AD), com o uso de linguagem documentária, e sua preocupação com metodologias de análise documentária e de construções de linguagens documentárias, são o complemento ideal para a abordagem da recuperação da informação, ajudando a solucionar parte dos problemas dos sistemas de informação.

2.4 - A recuperação e a representação da informação estatística

A preocupação por parte das instituições produtoras de informação estatística em disseminar os resultados dos seus trabalhos é bastante antiga, o que pode ser comprovado pela existência de publicações que já completaram mais de um século de existência e que veiculam essas informações. Mas a discussão que se coloca no momento não está focada no ato simples de tornar públicas as informações. É preciso responder uma série de indagações:

“Quem utiliza as informações disponibilizadas? Com que facilidade e nível de compreensão? Qual é o significado dessas informações para o público? Qual é a relação entre as informações divulgadas e as demandas informacionais da população? Que aplicação é feita dessas informações? (...) Perguntas do gênero devem ser feitas para que se possa aquilatar a extensão das realizações concretas de disseminação, e muitas delas são válidas mesmo nas situações em que são utilizados meios tradicionais de disseminação” (Lara & Conti, 2003, p. 27).

“A rigor não existe o usuário da informação pública governamental, mas os de diferentes motivações, origem, nível de instrução e idade. (...) alguns buscam informação utilitária, relacionada à satisfação de necessidades básicas; outros são motivados pela necessidade de conhecer para manter sua sobrevivência em determinados grupos; outros ainda procuram por informação por vontade de auto realização” (Lara et al. 2002, apud Lara & Conti, 2003, p.29).

Questões que põem em relevo o usuário, também são apontadas por Senra.

“De fato, uma vez que cada usuário perfaz um universo distinto, com desejos e com necessidades particulares e específicas, o sucesso da disseminação implicará mesmo o intenso promover e estimular do auto-atendimento, o que não impede de haver sempre espaço para um atendimento padronizado (na forma de produtos e serviços ajustados a uma tipologia básica de usuários) e um atendimento personalizado (na forma de produtos e serviços ajustados aos desejos de alguns usuários, sob a assessoria de especialistas nas estatísticas, conhecedores de suas metodologias)” (Senra, 2002, p. 80).

Mesmo enquanto essas questões estão sendo debatidas, já é possível prever que a forma de disponibilização da informação estatística pode ser melhorada, se usados os recursos tecnológicos disponíveis, para facilitar seu acesso.

Como bem atestam Guizzardi & Conti (2001) e Jannuzzi & Gracioso (2002), até a década de 1990 a publicação impressa era a forma preponderante de disseminação da informação estatística. O que se pode observar é que a esse gênero de publicação corresponde uma forma de organização das informações. Nessas publicações, a forma mais comum de organização dos dados para divulgação é a divisão das informações em temas/capítulos que mais ou menos reproduzem a organização de conhecimento tal como é encontrada na produção dos dados e na forma de organização institucional do órgão que a produz. Numa publicação desse tipo, uma pergunta sobre “quantas pessoas morreram de tuberculose no Estado de São Paulo” poderá ser respondida pelas informações do capítulo/tema Demografia⁶, mas muitas vezes essa informação não é encontrada pelo usuário que a procura no capítulo Saúde. Isso sem contar que a informação está inserida numa tabela, muitas vezes com título genérico, onde a palavra Tuberculose não aparece.

Para melhor entendimento dessa questão, reproduzimos como exemplo, parte de uma tabela extraída do Anuário Estatístico de São Paulo, disponível no *sítio* da Fundação Seade. Como se pode ver nessa tabela, encontrar a resposta para a pergunta efetuada exige do usuário algum conhecimento prévio sobre a produção do dado (para localizar o capítulo) e sobre a Classificação Internacional de Doenças (para localizar a causa de morte dentro da tabela, que, completa, possui 12 páginas).

⁶ Essa informação é produzida, na Fundação Seade, através dos registros de óbitos, de responsabilidade do setor de pesquisa demográfica.

Demografia: Mortalidade

25- Óbitos Gerais, por Grupos de Idade, segundo Causas de Morte Condensadas e Sexo Estado de São Paulo 2001

Causas de Morte e Sexo	0 a 4 Anos	5 a 14 Anos	15 a 44 Anos	45 a 64 Anos	65 Anos e Mais	Total (1)
TOTAL	12.029	2.070	45.556	59.046	114.388	234.073
Homens	6.597	1.273	35.319	38.003	57.318	139.398
Mulheres	5.432	797	10.237	21.043	57.070	94.675
1-001. Algumas Doenças Infecciosas e Parasitárias (A00-B99)	805	108	4.280	2.875	2.634	10.717
Homens	440	60	2.936	1.982	1.338	6.768
Mulheres	365	48	1.344	893	1.296	3.949
1-002. Cólera (A00)	-	-	-	-	-	-
Homens	-	-	-	-	-	-
Mulheres	-	-	-	-	-	-
1-003. Diarréia e Gastroenterite de Origem Infecciosa Presumível (A09)	277	3	42	86	307	716
Homens	152	1	24	47	123	348
Mulheres	125	2	18	39	184	368
1-004. Outras Doenças Infecciosas Intestinais (A01-A08)	29	-	5	5	9	48
Homens	17	-	3	4	3	27
Mulheres	12	-	2	1	6	21
1-005. Tuberculose Respiratória (A15-A16)	6	2	378	481	277	1.152
Homens	5	1	291	392	198	894
Mulheres	1	1	87	89	79	258
1-006. Outras Tuberculoses (A17-A19)	4	1	78	50	31	164

Fonte: Fundação Sistema Estadual de Análise de Dados – Seade.

(1) Inclui idade ignorada.

Nota: A partir de 1996, o agrupamento de causas de morte refere-se à X Revisão da Classificação Internacional de Doenças, desta forma, comparações com dados relativos a anos anteriores a esta data devem ser feitas com ressalvas.

Resumindo, a localização de uma informação estatística não é uma tarefa para pessoas que não tenham ao menos alguma vivência na sua produção e uso.

No final da década de 1990, a Internet transformou-se no meio principal de disseminação, mas de uma forma geral transpôs-se para o ambiente *web* a publicação tal qual era no papel.

Mesmo nos casos em que a informação foi reformatada para se adequar ao novo ambiente, a forma anterior de organização das informações foi mantida, permanecendo, assim, todas as dificuldades de localização da informação e acrescentando-se um sério agravante: o usuário da Internet não conta com um mediador para auxiliá-lo na busca da informação. A isso, soma-se o fato das instituições produtoras de informações estatísticas não terem a prática de organizar estatística para o público leigo.

Essa percepção é clara na afirmação de Guizzardi & Conti:

“... ser iniciado no assunto de que tratam os dados é ainda uma vantagem fundamental, já que as variáveis estão agrupadas em temas, às vezes em sub-temas e quase nunca em assuntos, dentro de tabelas em produtos com o nome das pesquisas sob as quais são construídos (Censo, PNAD, PED, PME, etc.). O leigo, aquele que não domina as especificidades e peculiaridades dos agrupamentos de dados ou desconhece o conteúdo dos produtos, tem de enfrentar a tarefa de navegar longamente pela página, tendo como aliados o bom senso e a sorte e como inimigos, além do azar, as armadilhas que, inconscientemente, é verdade, lhes prepara o subjetivismo de quem organiza a página” (Guizzardi & Conti, 2001, p. 53).

Além das dificuldades advindas da organização da informação, existe também aquela originada da prática de disponibilizar informação estatística em tabelas, adotada há mais de um século. Mesmo contando hoje com os recursos

trazidos pela informática, essa prática persiste sem que os responsáveis pela formatação da informação percebam o quanto esta forma dificulta a busca e a recuperação da informação.

Para facilitar o entendimento dessa colocação, podemos exemplificar com o produto 'Município de São Paulo – MSP', lançado pela Fundação Seade, em cujo lançamento se destacou o fato de ser um produto que disponibiliza informações desagregadas por subprefeituras:

“Além de agregar um conjunto de dados e indicadores socioeconômicos dispersos em variadas mídias, este MSP apresenta praticamente a totalidade deles desagregados por subprefeituras. Sem perder de vista os diferenciais regionais que marcam o Município, busca-se oferecer ao usuário o acesso mais rápido e fácil às informações desejadas” (<http://www.seade.gov.br/produtos/msp/>).

Apesar de ser verdadeiro o fato de que as informações estão desagregadas por subprefeitura, o mesmo não verifica no que diz respeito ao acesso mais rápido e fácil às informações desejadas. Uma vez que o produto foi concebido em forma de tabelas, qualquer usuário que quiser recuperar o perfil de uma determinada subprefeitura vai precisar consultar as quase quinhentas tabelas que compõem o produto, anotar a informação que cada uma delas traz sobre a subprefeitura em estudo e, só depois desse demorado, oneroso e complicado processo é que poderá obter a informação demandada.

No entanto, se o produto tivesse sido concebido como um banco de dados, todo esse processo seria realizado com a simples seleção da região desejada, operação impossível de ser realizada quando as informações estão confinadas às grades (se permitem o trocadilho) de uma tabela.

Num levantamento não exaustivo de *sites* de instituições produtoras de estatísticas, pudemos constatar que essas afirmações sobre a forma de disponibilização desses dados na Internet se confirmam, como se pode observar no quadro a seguir

Quadro 5 – Tipos de disponibilização de informação estatística

ESTADO	ÓRGÃO	TIPO
Acre	SEPLANDS	1
Bahia	SEI	1 e 2
Ceará	IPLANCE	1 e X
Espírito Santo	Instituto Jones dos Santos Neves	2
Mato Grosso	Fundação de Pesquisas Cândido Rondon	1
Mato Grosso do Sul	Seplanct	1
Minas Gerais	Fundação João Pinheiro	1
Pará	Secretaria Executiva de Planejamento e Coordenação Geral	2
Paraná	IPARDES	2 e X
Pernambuco	CONDEPE/FIDEM	1 e 2
Rio de Janeiro	CIDE	1 e 3
Rio Grande do Norte	IDEMA	1
Rio Grande do Sul	FEE	2
Santa Catarina	Secretaria de Estado do Planejamento	2
Sergipe	SEPLANTEC	1
Brasil	FIBGE	1, 2, 3

Tipo de disponibilização:

- 1 - Reprodução de Publicação Impressa
- 2 - Consulta a banco de tabelas
- 3 - Consulta a banco de dados
- X - Banco de dados não está acessível

Como pode ser visto no quadro acima, dos 16 *sites* consultados, 11 disponibilizam informação reproduzindo publicações impressas. Destes 11, 6 só disponibilizam informação desta forma. O *site* do Ceará anuncia um banco de dados, que em cinco tentativas feitas em dias diferentes, nunca foi possível acessar, o que eleva (nesse caso) para 7 o número de *sites* que só reproduzem publicações.

Nesse levantamento foi possível constatar que a disponibilização de informação estatística em banco de dados ainda é muito incipiente. No caso do Paraná, o banco só está disponível em uma rede estadual.

O *site* Fundação Seade não foi incluído nessa tabela, por ser o objeto de estudo desse trabalho. De qualquer forma, é um *site* que disponibiliza informação estatística de todas as formas.

Independentemente da forma com que os *sites* disponibilizam a informação, constata-se que:

“Não é raro verificar que os usuários encontram dificuldades para localizar informações, pois na construção do site e das páginas não se considera se ele conhece ou não as principais pesquisas da instituição ou não se questiona se eles entendem a linguagem na qual as informações são veiculadas” (Lara & Conti, 2003, p. 32)

O uso de uma linguagem de organização de informações seria um auxiliar para a comunicação entre especialistas e público leigo. Naturalmente a LD sozinha não pode resolver todos os problemas de recuperação de informação. Com a informação estatística, entretanto, trabalhamos a partir da hipótese de que a representação e a recuperação podem ser melhor solucionadas a partir do uso das metodologias de construção de Linguagens Documentárias e das metodologias da Terminologia.

O entendimento de que a informação estatística representa um contexto que não está explícito no dado, é primordial para pôr em relevo um aspecto importante que permite identificar as principais diferenças entre a representação documentária de informações estatísticas e a representação de informações textuais.

Para se representar conteúdos de textos, utilizam-se linguagens documentárias, destacando-se que a representação é de natureza referencial, aludindo aos assuntos tratados nos documentos; as representações não têm a função de substituí-los. As representações são, desse modo, generalizantes, sendo que o maior ou menor grau de especificidade do assunto a ser tratado depende do maior ou menor grau de reprodução de relações presentes na Linguagem Natural e no universo do conhecimento que pretendem cobrir.

A representação do discurso textual pode ser generalizante e de natureza referencial porque o contexto do termo utilizado para representá-lo está explícito no próprio texto. A função da linguagem documentária é encaminhar o usuário aos documentos sobre o assunto que o interessa.

Quando trabalhamos com a representação documentária de dados estatísticos, a informação a ser recuperada é numérica. A tradução do significado desse dado numérico em palavras mostra-se como uma operação delicada à medida que esses números já são uma representação específica de uma realidade, estabelecida em pressupostos que nem sempre conseguimos reconhecer e cujo significado também se altera conforme o contexto em que foram construídos.

O dado estatístico, por si só, não explicita o seu contexto, o que nos leva à necessidade de criar mecanismos que o façam, mostrando o recorte do universo temporal e social em que ele foi construído. Isso nos encaminha a levantar a hipótese de que a representação documentária da informação estatística não deve ser generalizante, pois seu conteúdo informacional é único.

Como já foi dito anteriormente, a informação estatística possui peculiaridades que não podem ser desprezadas pelo processo documentário. Com o intuito de facilitar a compreensão dessas características, achamos conveniente agrupar, aqui, algumas definições importantes:

“As estatísticas públicas correspondem ao dado social na sua forma bruta, não inteiramente contextualizado em uma Teoria Social ou uma Finalidade Programática, só parcialmente preparado para uso na interpretação empírica da realidade”. (...) “As estatísticas públicas – dados censitários, estimativas amostrais e registros administrativos – constituem-se, pois, na matéria prima para a construção de indicadores sociais” (Jannuzzi, 2001, p. 16).

“Indicador social é uma medida em geral quantitativa dotada de significado social substantivo, usado para substituir, quantificar ou operacionalizar um conceito social abstrato, de interesse teórico (para pesquisa acadêmica) ou pragmático (para formulação de políticas). É um recurso metodológico, empiricamente referido, que informa algo sobre um aspecto da realidade social ou sobre mudanças que estão se processando na mesma” (Jannuzzi, 2001, p.15).

“Microdados consistem no menor nível de desagregação dos dados de uma pesquisa, retratando, na forma de códigos numéricos, o conteúdo dos questionários, preservado o sigilo das informações. Os microdados possibilitam aos usuários, com conhecimento de linguagens de programação ou softwares de cálculo, criar suas próprias tabelas de dados numéricos. Acompanham os arquivos de microdados uma documentação que fornece nomes e respectivos códigos das variáveis e suas categorias, adicionada, quando necessário, dos elementos para o cálculo dos erros amostrais” (Koucher, 2002, p. 47).

Diante desse quadro, na tentativa de organizar a informação estatística, começamos por identificar suas características mais marcantes, o que pode orientar a elaboração de um sistema de recuperação desse tipo de informação.

Um dado estatístico é composto dos seguintes elementos:

- um fenômeno passível de ser quantificado (a variável);
- uma localização geográfica;
- um período temporal.

Chamamos genericamente de variável todas as medidas encontradas nas diversas pesquisas ou consolidadas através dos inúmeros registros administrativos, e também por extensão, indicadores ou índices resultantes das relações entre essas medidas.

“O termo *variável* tem uma história ambígua. Originário da matemática e da física teórica, o conceito de variável foi apropriado pelas ciências sociais, tomando um sentido cada vez mais amplo. Pouco a pouco foi estabelecendo-se a idéia ou o sentido de compreender, no conceito de *variável*, o resultado da separação de conjuntos de objetos selecionados, conforme um ou vários critérios específicos, em função de características determinadas, como, por exemplo, sexo, idade etc” (Dencker & Da Viá, 2001, p.127).

Em outras palavras, variável é um fenômeno medido, que ganha sentido quando localizado no tempo e no espaço, o que lhe confere uma dimensão. A informação estatística não existe fora dessa tríade. Sempre é composta por um acontecimento mensurável, num determinado lugar, em determinado período.

Os componentes tempo e espaço podem ser utilizados no sistema de recuperação de informação como chaves de acesso, sem grandes problemas, uma vez que esses atributos da informação já são devidamente normalizados e relacionados às variáveis correspondentes no próprio sistema de armazenamento dos dados, eliminando a necessidade de indexá-los através de um vocabulário.

Se a indexação dos componentes tempo e espaço é resolvida sem muitas dificuldades, o mesmo não acontece com os assuntos, tarefa que, como veremos no capítulo 4, é bastante complexa.

3 – A Construção de Linguagem Documentária – aspectos teóricos

Esta pesquisa baseia-se em dois objetos distintos: o dado estatístico e a linguagem documentária. A união desses dois objetos numa mesma perspectiva de estudo nasceu da visão que vimos adquirindo, ao longo da nossa vida profissional, da importância da informação estatística para o uso do cidadão comum, da observação de que seu uso ainda é bastante restrito aos setores governamentais e das dificuldades de disseminação dessa informação.

A procura de um caminho que pudesse esclarecer as incertezas acumuladas e respondesse ao anseio de compreensão do próprio objeto de trabalho, começou a ser elucidada por Barreto, que para instrumentalizar seu trabalho descrito em um artigo sobre a oferta e a demanda da informação usou o seguinte conceito de informação:

“Conjuntos significativos com a competência e a intenção de gerar conhecimento no indivíduo, em seu grupo, ou a sociedade” (Barreto, 1999, p. 1).

Barreto, no mesmo estudo, ainda acrescenta:

“Assim definido o conceito, a informação fica qualificada como um instrumento modificador da consciência do homem, e de seu grupo social. Deixa de ser, unicamente, uma medida de organização por redução de incerteza, para ser a própria organização em si” (Barreto, 1999, p. 1).

Smit & Barreto, 2002, ampliam este conceito de informação entendido como gerador de conhecimento, afirmando que, para que a informação seja investida da “competência de gerar conhecimento”, deve ter algumas características:

- estar explicitada num sistema significativo socialmente partilhado;
- deve ser decodificável;
- deve ser registrada em algum tipo de suporte.

A definição ampliada fica então da seguinte forma:

“Informação – estruturas simbolicamente significantes, codificadas de forma socialmente decodificável e registradas (para garantir permanência no tempo e portabilidade no espaço) e que apresentam a competência de gerar conhecimento para o indivíduo e para o seu meio. Essas estruturas significantes são estocadas em função de um uso futuro, causando a institucionalização da informação” (Smit & Barreto, 2002, p. 21-22).

Depois de assimilada essa noção de informação, o desafio de conseguir organizar a informação estatística para que possa ser transmitida, apreendida e usada pelo maior número de pessoas se volta à busca de parâmetros para a consecução desses objetivos.

No contexto atual, a informação estatística, que é o objeto de disseminação, é normalmente disponibilizada na Internet, portanto, como já mencionamos, sem as mediações presenciais que sempre permitiram alguma forma de auxílio na busca e seleção das informações para responder às indagações dos usuários não especializados. Esse novo ambiente (ou suporte, ou meio de transmissão) tornou essa tarefa não só mais complexa como também mais imperativa, à medida que cada vez mais a informação está sendo procurada pelo usuário remoto.

3.1 – Referencial Metodológico de Construção de Linguagem Documentária

Nas pesquisas sobre Análise e Linguagem Documentária, buscamos os referenciais metodológicos para consolidar e expandir conhecimentos que instrumentalizem os objetivos de construir canais de comunicação da informação estatística, de forma a possibilitar que ela seja apreendida pelo cidadão comum, possibilitando o reuso dessa informação.

Vale repetir:

“A função da LD é tratar o conhecimento dispondo-o como informação. Em outras palavras, compete às LDs transformar estoques de conhecimento em informações adequadas aos diferentes segmentos sociais. É esse partilhamento que está na base do caráter público da informação e que não pode ser obtido na ausência de uma LD” (Cintra et al. 2002, p.16-17).

“A partir do enfrentamento das várias linguagens - a dos especialistas, a dos utilizadores de informação -, são modelizadas as linguagens documentárias, instrumentos que permitem a materialização da representação de informações, de forma a garantir a comunicação. As linguagens documentárias funcionam, assim, como meio de estabelecer as pontes que viabilizam o acesso e uso da informação” (Lara, 1998, p.100).

A tarefa de representar convenientemente as informações estatísticas e disponibilizá-las para uso da sociedade, levou-nos a universos de estudos fundamentais para a identificação, não só da necessidade de dispormos de um vocabulário próprio para essa tarefa, como também de compreender a sua complexidade, principalmente no aspecto que se relaciona ao controle semântico do seu vocabulário, bem como os meios para sua manutenção e uso.

“Para a Análise Documentária, a noção de representação envolve a construção de ‘substitutos’ de textos, que por uma razão econômica devem ser apresentados de forma compacta para poderem integrar sistema de informações. A natureza metodológica da Análise Documentária permite, entretanto, transcender a noção de documento ‘textual’, englobando a representação da informação independentemente da forma de registro e apresentação das informações (informação estatística, informação pública institucional; papel, cds, documentos eletrônicos, etc)” (Lara, 1999, p.8).

A análise documentária compreende necessariamente os processos de indexação e de recuperação. O conjunto completo de termos utilizados como referência em um sistema de indexação/recuperação é denominado vocabulário ou linguagem de indexação daquele sistema, designações que remetem ao conceito de linguagem documentária.

A indexação de documentos via linguagem documentária se estabelece principalmente em duas operações:

1 - análise conceitual - envolve a leitura e a síntese do texto e a decisão sobre o assunto de que trata o documento.

2 - tradução ou representação - envolve a seleção de termos de uma linguagem documentária para representar o conteúdo dos documentos. Podemos nos referir a eles como termos de indexação ou descritores, termos normalmente utilizados quando o instrumento de representação usado é um tesauro.

Quando falamos de recuperação de informação, o processo que ocorre é semelhante ao do processo de indexação:

1 - análise conceitual - diz respeito à interpretação das perguntas do usuário e sua expressão lingüística.

2 - tradução - diz respeito à representação da necessidade do usuário através dos termos do sistema, o que constitui um elemento importante da estratégia de busca que recupera a informação.

Em esquema, temos:

Fig.1 – Uso do vocabulário na indexação e na consulta

Como demonstra a figura 1, o vocabulário empregado em um sistema de informação é da maior importância, pois os assuntos, tanto das consultas quanto dos documentos devem ser representados por termos tomados dessa linguagem.

A construção de índices com base em linguagens documentárias, pode ser definida como uma

“Representação por tradução: formulação em linguagem que não se confunde com a linguagem do texto, mesmo que os termos apresentem a mesma forma” (Lara, apontamentos de aula, 2003).

Na maioria dos sistemas de recuperação o vocabulário empregado será um vocabulário controlado, o que chamamos de linguagem documentária. A maior parte das linguagens documentárias está organizada de maneira lógico-semântica. A diferença básica entre os sistemas de classificação bibliográfica mais utilizados (CDD/CDU) e os tesouros está no fato de os primeiros estarem organizados a partir de dez classes de assunto pré-determinados, baseados em convenções culturais e ideológicas consensuais à época em que foram desenvolvidos (final do séc. XIX).

Os tesouros, por sua vez, estruturam o conhecimento a partir de temas ou categorias em função dos objetivos específicos do sistema de informação para os quais são elaborados. Os tesouros podem ou não adotar facetas para promover o sub-agrupamento dos termos em uma categoria, permitindo assim, contemplar simultaneamente vários pontos de vista sobre um conceito. Sua vantagem está na flexibilidade que permite construir tantas organizações de universo nocional quanto forem os pontos de vista contemplados pelas diferentes disciplinas e necessidades.

O tesouro é uma linguagem documentária formalmente organizada de modo a explicitar as relações conceituais entre seus termos. É apropriada para a indexação pós-coordenada, ou seja, aquela em que as combinações entre os termos descritores são feitas no momento da recuperação da informação. Para a elaboração de um vocabulário, utilizando os parâmetros encontrados em “Diretrizes para o

estabelecimento e desenvolvimento de Tesouros Monolíngües” (AUSTIN, 1993), é preciso distinguir bem as relações básicas que unem seus termos.

No quadro a seguir, procuramos sintetizar essas relações, demonstrando seus significados, usos e representações.

Quadro 6 – Relações básicas de um vocabulário controlado

Tipo de Relação	Definição	Finalidade do Uso	Representação
Equivalência	Relação entre o termo preferido e o não preferido, utilizado quando dois ou mais termos são considerados, para fins de indexação, como referentes ao mesmo conceito.	Indicar qual o termo preferido ou preterido pelo sistema no caso de sinônimos, quase sinônimos ou mesmo no caso de deslocamentos genéricos.	USE (indica o termo preferido) UP (usado para - indica os termos considerados equivalentes)
Hierárquica	Relação baseada em graus ou níveis de superordenação/subordinação: Gênero/Espécie ou relação genérica, ou baseada em níveis de ordenação: Todo/Parte ou relação partitiva.	Indicar uma noção fundamental que inclui noções específicas. Desse modo, as relações genéricas/partitivas indicam que todo conceito específico é parte de um conceito amplo.	TG (termo genérico) TE (termo específico)
Associativa	Relação baseada em associações de sentido, num determinado contexto. A ocorrência e utilidade de qualquer associação dependem do modo de organização dos domínios de especialidade e das necessidades do usuário	Indicar pares de termos que não pertencem à mesma hierarquia mas que se inter-relacionam. Devem ser utilizados quando a consulta às informações indexadas pelo outro termo complementarem a pesquisa.	TR (termo relacionado)

Além do resumo apresentado nesse quadro, podem-se destacar alguns outros recursos, como:

a) Deslocamento genérico – se refere a termos considerados demasiado específicos para adoção como termos de indexação, mas que podem ser mapeados com os termos mais gerais e mais adequados ao sistema.

Ex.:

Receita Municipal Corrente - 45.01.01
 TT Finanças Públicas Municipais
 UP Receita Municipal por Multas e Juros de Mora
 Receita Municipal por Indenizações e Restituições
 Receita Municipal da Dívida Ativa
 Receitas Municipais Diversas

b) Notas explicativas - quando a aplicação de um termo puder gerar alguma ambigüidade, ele pode receber uma nota explicativa identificada com o prefixo **NE** no vocabulário. Pode ser uma verdadeira definição de dicionário, pode indicar como um termo deve ser interpretado no contexto daquele vocabulário e/ou distingui-lo de outros termos com os quais possa ser confundido.

Ex.: Documentação

NE: processo de armazenar e recuperar informação em todos os campos do conhecimento.

c) Definições – a partir das referências da Terminologia, podemos fazer uma distinção entre as notas explicativas e as definições: as notas são mais adequadas para se referirem a questões de aplicação do termo, já as definições devem ser usadas para delimitação da intensão/extensão do termo utilizado.

Ex.: Investimentos Municipais Per Capita

Definição: Valores referentes a Investimentos que englobam Gastos em Obras e Instalações, Equipamentos e Material Permanente, Dotações para Constituição ou Aumento do Capital de Empresas que não sejam de carácter comercial ou financeiro.
 Este indicador demonstra o quanto cada município despendeu por habitante, durante os exercícios especificados, para estas categorias de gastos.

A partir do momento que incluimos, no nosso sistema de linguagem controlada, um recurso como o do uso de definições, passamos a construir um vocabulário com contribuições oriundas da Terminologia, cujos parâmetros veremos a seguir.

3.2 - Referencial Metodológico da Terminologia

Na construção de uma Linguagem Documentária, deve-se ter sempre presente o fato de que o rigor utilizado na sua elaboração é que vai fornecer a garantia de um instrumento que é “simultaneamente um modo de organização e uma forma de comunicação” (Tálamo, 1997).

Focando-se a função de comunicação, é importante destacar que:

“a linguagem documentária não se traduz como instância mediadora pela simples presença de expressões retiradas dos documentos: ao contrário, para funcionar efetivamente como linguagem é preciso que suas unidades estejam organizadas de forma a configurar determinados significados” (Lara, 2002, p.2).

É nessa configuração de significados que encontramos a segunda característica da linguagem documentária: a de organizadora de informações.

Para a construção da LD, é preciso articular três aspectos diferentes: a língua em que o sistema é construído, pois é este o instrumento de comunicação; a língua da especialidade, pois é nela que a informação é construída; o fundamento originário da Ciência da Informação, que se caracteriza como:

“... o modo próprio como são efetuados os recortes nas áreas de especialidade e na linguagem natural, transformando as unidades desses dois códigos em unidades de informação expressos em linguagem documentária: a delimitação de tais unidades depende de objetivos institucionais, onde se incluem as necessidades do usuário para qual o sistema de dirige” (...) “a linguagem documentária deve simular, através de uma rede de relações entre os seus termos, a representação de um universo informacional de uma área de especialidade ou atividade, útil e passível de ser reconhecido (ou conhecido) pela comunidade de usuários a que se destina” (Lara, 2002 p. 3).

Na primeira parte deste capítulo, resumimos os principais parâmetros dados pela norma ISO 2788:1986/BS 5723:1987, que correspondem às Diretrizes para o estabelecimento e desenvolvimento de tesauros monolíngües (Austin, D. & Dale, P., 1993). Para correlacioná-los aos princípios terminológicos de organização de sistemas de conceitos, resumimos, aqui, parâmetros da Terminologia. A associação dos dois conjuntos de princípios permite conferir maior rigor ao desenvolvimento de linguagens documentárias.

Acreditamos que a primeira noção da terminologia que devemos considerar é a de “conceito”, descrita pela norma ISO 1087-1:2000 como unidade de conhecimento constituída de:

- a) síntese do objeto (elemento da realidade passível de ser percebido ou concebido);
- b) predicados necessários a ele atribuídos;
- c) signos lingüísticos para fins de comunicação.

Consideramos que a apreensão desta noção é o primeiro e mais fundamental passo para a construção de linguagens documentárias: trabalhamos com unidades de informação e sem o reconhecimento de que organizamos conceitos ou significados, não podemos determinar corretamente quais signos são seus representantes. Da mesma forma, sem o conhecimento das características a eles atribuídas, não os definimos, como também não construímos redes de relações consistentes, o que significa não cumprir o papel de possibilitar a organização e a comunicação documentária.

Levando-se em consideração que a linguagem documentária tem a característica de intermediação entre o sistema de informação e os usuários, devemos prever formas de explicitar as noções envolvidas em cada unidade de

informação e suas relações, para tentar garantir ao indexador, usuário desse instrumento, e ao receptor, destinatário final do sistema de informações, as mesmas condições para compreender a lógica utilizada na construção da linguagem documentária. Essa necessidade nos leva à segunda apreensão importante dos instrumentos terminológicos: o estabelecimento das relações entre os conceitos.

Os conceitos são delimitados a partir das características a eles atribuídos. Uma característica é um aspecto a partir do qual se realiza a representação de um objeto.

Ex: uma característica de **Atividades Bancárias** é ser uma atividade exercida por **Bancos**. Uma característica de **Contas de poupança** é ser uma das ações promovidas pelos Bancos. Por sua vez, uma das características de um **Banco** é ser uma **Instituição Financeira**. A partir da identificação dessas características, **contas de poupança** é um termo conceitualmente relacionado a **atividades bancárias**, a **bancos** e a **instituições financeiras**.

A identificação de características permite delimitar os conceitos e através de suas oposições e similitudes, as relações entre eles.

As relações entre os conceitos retratam um sistema de conceitos, composto principalmente de:

- a) conceito genérico;
- b) conceito específico;
- c) conceito partitivo;
- d) conceito coordenado;
- e) conceito associado.

“Os sistemas de conceitos variam em sua organização, uma vez que diferentes abordagens dos domínios são possíveis” (Lara, 2004, p.238).

Uma vez determinadas as relações entre os conceitos, chegamos à terceira apreensão importante: a elaboração de uma definição.

A definição é, segundo a norma ISO 1087-1:2000, um:

“Enunciado que descreve um conceito, permitindo diferenciá-lo nitidamente dos demais conceitos dentro do sistema”.

Na realidade, o processo de identificação de características dos conceitos e do estabelecimento das relações entre eles, faz parte do processo definicional. A fase ora especificada compreende o processo de sua expressão lingüística, de forma a facilitar a compreensão do conceito.

O uso regular de definições é o instrumento que garante a explicitação da lógica da construção da linguagem documentária, apontada como necessária em parágrafo anterior.

“A definição é por excelência, classificadora, hierarquizante, estruturante” (Desmet, 1990, *apud* Lara, 2004 p. 238).

As definições compreendem vários tipos. As normas terminológicas destacam as definições intensionais e as extensionais.

“As definições intensionais devem basear-se nas relações conceituais (genéricas ou partitivas) determinadas na análise e são preferíveis a outros tipos de definições, uma vez que mostram mais claramente as características essenciais do conceito no interior de um sistema de conceitos (ISO 704:2000). A definição intensional feita a partir de relações genéricas corresponde à definição aristotélica: gênero próximo, diferença específica. É iniciada pela menção do conceito genérico seguido das características distintivas do conceito sob definição” (Lara, 2002 p. 13).

Exs.:

Índice de Apgar 1º minuto

Definição: Medida de vitalidade da criança, realizada no 1º minuto de vida, através do teste de Apgar, que avalia, no 1º e no 5º minuto de vida, os seguintes itens: frequência cardíaca, respiração, tônus muscular, cor da pele e resposta a estímulos nervosos.

Consultas de Pré-natal

Definição: Número de consultas de pré-natal que a mãe freqüentou durante a gestação. A assistência pré-natal é um conjunto de cuidados médicos, nutricionais, psicológicos e sociais, efetuados através de consultas médicas periódicas, destinados a proteger a criança e a mãe durante a gravidez, parto e puerpério. Os números de consultas são agrupados conforme a Declaração de Nascimento.

“A definição extensional (ou definição por extensão) descreve o conceito enumerando todos os conceitos subordinados que correspondem a critério de subdivisão. A definição extensional é útil em domínios altamente especializados, quando é possível listar os objetos que fazem parte da extensão do conceito” (LARA, 2002, p.13).

Exs.:

Receita Municipal

Definição: Receitas orçamentárias recolhidas aos cofres públicos municipais por força de arrecadação, recolhimento e recebimento. Englobam as Receitas Correntes e as Receitas de Capital.

Tipo de Parto

Definição: Identificação do tipo de parto, que pode ser cesáreo ou vaginal, sendo que essa última alternativa inclui o parto por fórceps ou vácuo-extrator.

Acreditamos serem esses os principais elementos da terminologia que, apreendidos no momento da elaboração da linguagem documentária, muito colaboram, tanto na etapa de construção como no uso da referida linguagem.

4 – Linguagem Documentária para o *site* da Fundação Seade

A inclusão de novas tecnologias ao nosso cotidiano e a expansão acelerada da Internet propiciaram uma facilidade muito grande na produção e disseminação de informação, que, aliadas ao baixo custo que a Internet oferece para que a informação seja amplamente divulgada, tornou-a a mídia preferida das agências de estatística.

Vale lembrar, no entanto, que o uso de uma boa mídia não significa por si só uma boa transmissão de informação. Não basta disponibilizar a informação, é necessário criar condições para a que a informação seja apreendida e transformada em conhecimento.

“A informação é fluxo e o conhecimento é estoque. Semelhante afirmação não deve ser submetida aos critérios de verdade porque tem somente um caráter operacional. Permite de fato visualizar que o conhecimento, em termos comunicacionais, é algo passivo e que a informação, vista sob o ângulo citado, trata de ativá-lo por meios de variados produtos informacionais. Enquanto fluxo, os produtos informacionais disseminam a informação porque determinam acessos específicos de modo que os públicos tenham condições reais de consumi-la. É preciso ainda observar, para esclarecer ainda melhor a associação feita de início, que não existe relação unívoca entre estoque e fluxo. Ao contrário, aceitando-se a idéia de que a produção da informação segue regra temática enquanto a do conhecimento científico segue orientação disciplinar, tem-se que os produtos informacionais encontram-se na dependência dessas fragmentações temáticas processadas. De modo geral a partir de um mesmo estoque pode-se gerar inúmeros produtos informacionais, isto é, fluxos socialmente determinados para consumo diversificados” (Tálamo, 2005, p.7).

De uma forma geral, as agências de estatística divulgam suas informações ou sob uma formatação muito rígida, ou ainda utilizando os mesmos princípios adotados na fase de produção dos dados. A forma mais freqüente é reproduzir seus produtos tradicionais que foram originalmente desenhados para uma mídia impressa. Não existem, ainda, grandes preocupações em adequar os produtos de

informação estatística à Internet e, muito menos, em criar produtos diversificados para atingir públicos diversificados.

“No campo de ação da disseminação das informações estatísticas, pode-se dizer que, de certo modo, as agências disseminam para seus pares, de modo que seus atuais usuários estão no mesmo estamento social, têm voz, podem realizar uma interlocução com os responsáveis pela oferta da informação. Como quem não tem voz, não reclama (ou mesmo que a tenha nem sempre possui os meios para fazê-lo), não é preciso muito esforço para identificar de onde advém a impressão de que o acesso às informações está universalizado, porque se entende que o ‘dever’ foi cumprido” (Lara & Conti, 2003, p. 31).

A alteração desse quadro demanda a criação de mecanismos para proporcionar acesso mais fácil e livre às informações, como também reclama a elaboração de produtos que sejam adequados a distintas necessidades de diferentes tipos de usuários. Do mesmo modo, é conveniente que tais mecanismos incluam a possibilidade de fornecer metainformações, de forma a facilitar a compreensão dos dados disponibilizados.

Foi com a preocupação de minorar os efeitos desses problemas, e de propiciar aos usuários um acesso mais fácil e livre às informações estatísticas, que a Fundação Seade inseriu, nos procedimentos de preparação dos seus produtos a serem divulgados na Internet, uma etapa de tratamento documentário, para que a disponibilização de seus dados se adequassem às necessidades de usuários distintos e para que eles pudessem contar com meios de acessar, muito rapidamente, informações complementares necessárias à compreensão da informação disponível.

A criação de instrumentos para a consecução desses objetivos é uma tarefa complexa, mas pode significar “um investimento formador e ampliador do universo dos públicos da informação estatística” (Lara & Conti, 2003, p.32). Trata-se de

investir esforços na criação de mecanismos que permitam organizar a informação para sua transferência e circulação mais ampla.

A realização dessa tarefa solicita mobilizar os conhecimentos referidos principalmente no capítulo anterior, tanto no que diz respeito ao desenvolvimento de uma linguagem documentária, quanto no que se refere à indexação das informações. No entanto, o desenvolvimento dessas tarefas não foi feito sem enfrentar inúmeras dificuldades, cujas características procuraremos relatar a seguir.

4.1 – A construção da Linguagem Documentária para dados estatísticos

No empreendimento de identificar princípios para elaborar um instrumento que represente e recupere dados estatísticos, verificamos que uma linguagem documentária com esse fim deve ser muito mais uma forma de organizar os dados eles mesmos para disponibilização do que, como ocorre na representação de textos, uma forma de prover sínteses dos textos representados. Isso acontece porque, como já dissemos anteriormente, essas informações – os dados estatísticos – são de natureza muito específica no seu recorte da realidade e, por conseguinte, sua representação deve ser realizada num nível mais particularizante do que aquele normalmente utilizado na representação de textos. Diferentemente do que ocorre com a representação de textos, qualquer tentativa de sintetizar as representações de dados estatísticos através, por exemplo, de pistas para recuperação, pode gerar expectativa de informação nem sempre passível de ser satisfeita.

A afirmação acima pode ser confirmada através de exemplos.

Ex.: “Óbitos por acidentes de trânsito”

Se desdobrarmos o enunciado acima e representarmos essa informação sob a rubrica “**Acidentes de trânsito**”, podemos, dentro de um sistema estatístico, gerar no usuário a expectativa de encontrar, como resposta, *o número de acidentes de trânsito ocorridos*, informação que não necessariamente existe no sistema. É necessário, portanto, representar um dado estatístico num maior nível de especificidade, como por exemplo, sob o termo “**Óbitos por acidentes de trânsito**”, que, nesse caso já é o próprio nome da variável. Cabe aqui a observação de que nem sempre o nome da variável é normalizado, o que às vezes também provoca

dificuldade no entendimento da informação. Aliás, esse é um outro aspecto que confirma a necessidade de utilizar os princípios metodológicos previstos para a construção de uma linguagem documentária.

No momento em que os termos utilizados no vocabulário se aproximam quase que literalmente dos nomes das variáveis, resta, para a linguagem documentária, a função de organização dessa informação num conjunto com outras informações que com ela se relacionam, tanto de forma hierárquica como associativa. Por essa razão, defendemos a hipótese de que, no que tange às linguagens de organização e representação da informação estatística, o nível de especificidade da linguagem coincide com o nível da especificidade do vocabulário (ou dos termos) utilizado para a indexação. Decorre daí que, diferentemente do que acontece com a indexação de textos, a indexação de dados estatísticos não pode ser generalizante.

A função de síntese das representações, porém, também ocorre no caso do tratamento das informações estatísticas, mas em situações diferentes. Ela só vai estar presente no momento em que organizarmos um conjunto de informações, muitas vezes classificadas em temas, por exemplo. Os temas (ou categorias) representam, pela sua própria natureza, um nível de generalização e abstração maior. Em outras palavras, sintetizamos a informação quando a reunimos com outras informações que, no seu conjunto, traduzem um recorte generalizante, como um cenário construído sob uma determinada ótica, como acontece, por exemplo, com os termos "**Condições de Vida**", "**Violência**" etc., onde se despreza a especificidade da informação para oferecer ao usuário uma síntese temática que tem seu significado num universo mais geral. Sob o aspecto terminológico, tais termos contemplam formas de denominação freqüentemente utilizadas por inúmeros usuários, como também por produtores de informação.

Nesse contexto, tanto a informação de “*Óbitos por acidente de trânsito*” como a informação de “*Ocorrências de homicídio*” podem ser encontradas reunidas genericamente sob o termo **Violência**, ou, no outro exemplo, “*Número de cômodos no domicílio*”, “*Domicílio com acesso a ligação de água*”, e “*Posse de geladeira*”, por **Condições de Vida**.

Mantendo bem clara a noção de que o trabalho documentário é por natureza institucional, e portanto, não deve pretender atingir o usuário individual, mas um usuário coletivo que é previsto pela instituição, elegemos do ponto de vista prático, para o desenvolvimento deste trabalho, dois focos:

a) foco na produção (emissão)

Sob este ponto de vista, deve-se procurar usar as categorias generalizantes de informação que são razoavelmente uniformes e compartilhadas pelas instituições estatísticas;

b) foco no usuário (recepção)

Sob esta perspectiva, deve-se procurar “acompanhar” as formas de solicitação, as dúvidas e perguntas feitas pelo usuário, para introduzir tais referências na organização das informações, seja via uso de equivalências, seja via uso de poli-hierarquias, recursos que possibilitam que a linguagem de recuperação se aproxime da linguagem natural utilizada pelo usuário não especializado.

Outro aspecto a ser considerado na elaboração de uma linguagem para o tratamento da informação estatística é o uso da relação de equivalência, que tem, muitas vezes, uma função pedagógica. O exemplo a seguir explicita bem essa função de aproximação de diferentes linguagens – a linguagem de especialidade e a

do entendimento comum, o que supera, em termos de qualidade para o usuário, o simples uso de sinônimos.

Ex: **Extorsão mediante seqüestro**, termo específico de **Crimes contra o Patrimônio**.

O usuário não especializado que deseja buscar a informação sobre o número de seqüestros que ocorreram num determinado ano e local, provavelmente vai fazer a sua pergunta por **seqüestro** e não por **extorsão**, muito menos por **Crimes contra o patrimônio**, já que o rigor da classificação dada pelo Código Penal⁷ não é necessariamente conhecido pelo usuário que, mais provavelmente, poderia procurar a informação sobre seqüestros, na falta desse termo específico, em **Crimes contra a pessoa**.

Nesse caso, em que o uso do recurso do deslocamento genérico, também previsto nas metodologias de construção de linguagem documentária, afastaria mais ainda a linguagem do especialista da linguagem do público leigo, é necessário procurar utilizar a forma de entendimento popular como equivalente à forma utilizada pelo especialista.

Procedendo assim, tanto o especialista quanto o leigo podem recuperar a informação sem a necessidade de intermediação de um especialista no assunto, que, no caso, já nem é estatístico, mas, jurídico.

Numa outra situação, identificamos a necessidade de usar o termo “**Crimes hediondos**”, que entre outros, compreende *Latrocínio, Seqüestro e Estupro*, crimes

⁷ Código Penal (Decreto-lei nº 2.848, de 07/12/1940), artigo 159.

que estão formalmente subordinados, segundo o Código Penal, a Crimes contra o patrimônio e Crimes contra os costumes, respectivamente⁸. Uma solução possível é o uso da poli-hierarquia. A estrutura da linguagem documentária, contemplando esses dois exemplos, ficaria como segue:

Justiça e Segurança – 60

TE Ocorrências Policiais
Inquéritos Policiais

Ocorrências Policiais - 60.01

TT Justiça e Segurança
TE Crimes
Contravenções Penais

Crimes - 60.01.01

TT Justiça e Segurança
TG Ocorrências Policiais
TE Crimes Contra a Pessoa
Crimes Contra o Patrimônio
Crimes Contra a Incolumidade Pública
Crimes Contra os Costumes
Crimes Hediondos

Crimes Contra a Pessoa - 60.01.01.01

TT Justiça e Segurança
TG Crimes
TE Homicídio
Lesão Corporal
Tentativa de Homicídio
TR Crimes Contra os Costumes

Crimes Contra o Patrimônio - 60.01.01.02

TT Justiça e Segurança
TG Crimes
TE Estelionato
Extorsão
Furto
Roubo
Latrocínio

Extorsão - 60.01.01.02.02

TT Justiça e Segurança
TG Crimes Contra o Patrimônio
TE Seqüestro

⁸ Código Penal (Decreto-lei nº 2.848, de 07/12/1940): art. 157, § 3º, *in fine*; art. 159; art. 213.

Seqüestro - 60.01.01.02.02.01

TT Justiça e Segurança
 UP Extorsão Mediante Seqüestro
 TG Crimes Contra o Patrimônio
 Crimes Hediondos

Latrocínio - 60.01.01.02.05

TT Justiça e Segurança
 UP Roubo Seguido de Morte
 TG Crimes Contra o Patrimônio
 Crimes Hediondos
 TR Homicídios

Crimes Contra a Incolumidade Pública - 60.01.01.03

TT Justiça e Segurança
 TG Crimes
 TE Tráfico de Drogas
 Uso de Drogas

Tráfico de Drogas - 60.01.01.03.01

TT Justiça e Segurança
 UP Tráfico de Entorpecentes
 TG Crimes Contra a Incolumidade Pública

Crimes contra os Costumes - 60.01.01.04

TT Justiça e Segurança
 TG Crimes
 TE Estupro
 Atentado Violento ao Pudor
 TR Crimes Contra a Pessoa

Estupro - 60.01.01.04.01

TT Justiça e Segurança
 TG Crimes contra os Costumes
 Crimes Hediondos

Crimes Hediondos - 60.01.01.05

TT Justiça e Segurança
 TG Crimes
 TE Latrocínio
 Estupro
 Seqüestro

Utilizando os recursos de equivalência e poli-hierarquia, é possível permitir a consulta por termos utilizados pela linguagem coloquial, e ao mesmo tempo classificar esses crimes na hierarquia adotada pela instituição produtora de estatísticas, que geralmente segue a estrutura utilizada pelo órgão responsável pelo

registro público que gera os dados, sem deixar de utilizar também um segundo prisma, sugerido pelas próprias reorganizações da área jurídica, que qualifica esses crimes também pelo grau do dano provocado na sociedade, o que se confirma através de consultas à legislação pertinente (Anexo 1).

A prática para elaboração de linguagens documentárias mostra que quando se trabalha num universo de informações restrito ou muito especializado, o uso da poli-hierarquia deve ser evitado, porque pode confundir o usuário e provocar perda de precisão na recuperação. No entanto, quando o universo de informações é amplo e reúne várias áreas de especialidade (como ocorre com o conjunto de informações sociais, econômicas, demográficas, culturais disponibilizadas simultaneamente), o uso da poli-hierarquia constitui um valioso instrumento, quer para o relacionamento entre conceitos, quer para a localização da informação, porque permite atender a usuários que têm interesses diversos e que, muitas vezes, não conhecem as particularidades da terminologia de cada domínio referenciado.

Por exemplo, quando um usuário faz uma pesquisa pelo termo **Mulheres**, pode-se supor que ele visa encontrar todas as informações existentes no sistema sobre o gênero feminino. Normalmente esse termo faz parte da estrutura de Características Pessoais, que reúne atributos como idade, estado civil, sexo, etc. aplicáveis a vários domínios da informação estatística. Naturalmente o usuário não está interessado nessa faceta, que é uma faceta auxiliar, criada na linguagem para reunir termos importantes na representação da informação estatística e que não se enquadram nos grandes temas utilizados.

A solução ao problema acima é o uso da poli-hierarquia, pois ela permite reunir sob o termo Mulheres, todos os aspectos relacionados ao gênero feminino, oferecendo aos usuários uma possibilidade de acesso ao conjunto de informações pertinentes.

Nesse caso, a estrutura pode se apresentar sob a forma abaixo:

Mulheres - 03.03.01

TT Característica Pessoais

UP Gênero Feminino

TG Sexo

TE População Feminina

Óbitos Femininos

Mortalidade de Mulheres em Idade Fértil

Ocupação da Mãe

Escolaridade da Mãe

Eleitores femininos

Gestação de adolescentes

Outro aspecto interessante da poli-hierarquia é a consolidação da idéia de rede de relacionamentos entre os termos e informações, como mostra o desenho a seguir:

Fig. 2 – Representação de relações poli-hierárquicas

Confirma-se, desse modo, que para o tratamento da informação estatística a aplicação dos princípios previstos pelas diretrizes de elaboração de tesouros deve prever as especificidades desse gênero de informações, não sendo conveniente utilizá-las da mesma forma que para as informações textuais.

4.2 – Uso da Linguagem Documentária no *site* da Fundação Seade

Além da estrutura de relações, que provê e assegura os significados e a interpretação dos termos, um sistema de recuperação de informação estatística deve contemplar o acesso a definições, pois mais do que dados, representamos conceitos. As definições, além de auxiliarem o usuário leigo na tarefa de seleção e recuperação das informações, asseguram a própria estrutura da organização dos dados.

Para exemplificar, no caso já citado do termo “**Seqüestro**”, a inserção de uma definição auxilia o usuário a compreender seu significado no contexto do sistema de informações.

Ex.: Seqüestro

Definição: Retenção de pessoa com o fim de obter, para si ou para outrem, qualquer vantagem, como condição ou preço do resgate.

Outro recurso que deve ser inserido no sistema de recuperação de dados estatísticos é a disponibilização de notas explicativas sobre o processo da produção dos dados. Isso porque, muitas vezes encontramos, por exemplo, problemas na fase de coleta de dados, que podem provocar incompatibilidades entre os totais apresentados para o Estado e a soma dos resultados para municípios.

Ex.: Total de Empregos Ocupados

Nota: As totalizações para Regiões de Governo, Administrativas, Região Metropolitana e Estado incluem as informações municipais não divulgadas devido a sigilo estabelecido pelo Ministério do Trabalho.

Do mesmo modo, pode ser necessário incluir notas sobre a comparabilidade de séries históricas, como por exemplo, aquela que alerta sobre uma mudança de denominação que nem sempre corresponde a uma mudança muito grande de conceito:

Ex.: Ensino Médio

Nota: A denominação Ensino Médio substituiu a de Ensino de 2º Grau, a partir da aprovação da Lei de Diretrizes e Bases da Educação Nacional - LDB 1996.

Outra informação importante a ser disponibilizada é a fonte dos dados, que não só garante a credibilidade das informações como, aliada a notas e definições, explicam eventuais divergências entre informações coletas por instituições diferentes.

Ex.: Receita Municipal Arrecadada

Fonte: Ministério da Fazenda. Secretaria do Tesouro Nacional.
Fundação Sistema Estadual de Análise de Dados - SEADE

A partir do exposto até o momento, podemos sintetizar elementos complementares para a contextualização, aos elencados no Quadro 6 (p. 51), a serem incluídos, sempre que necessário, na apresentação dos dados estatísticos

No caso concreto da indexação dos dados para o *site* da Fundação Seade, encontramos algumas particularidades que, podemos dizer, são típicas dos dados estatísticos; outras dizem respeito ao sistema de recuperação que foi construído para disponibilizar essas informações. Vamos tentar separar essas particularidades, para melhor entendimento.

Dentre as armadilhas da indexação de dados estatísticos, a mais exemplar é a da indexação generalizante, como se pode verificar a seguir.

Indexando duas variáveis: “Operações de Crédito” e “Operações Municipais de Crédito”, a tendência do indexador apressado é reuni-las sob o termo **Operações de Crédito**. Como se pode observar nas definições que se seguem, as variáveis “Operações de Crédito” e “Operações Municipais de Crédito” fazem parte de contextos diferentes. A primeira é uma informação relativa ao sistema financeiro e retrata o montante de dinheiro que foi colocado no mercado a título de empréstimos; a segunda se refere ao setor público, mais especificamente, às finanças públicas municipais e retrata o montante arrecadado pela administração através de empréstimos obtidos e também de títulos públicos colocados no mercado.

Operações de Crédito

Definição: Empréstimos efetuados ao setor público e privado pelos bancos comerciais, múltiplos e caixas econômicas.

Operações Municipais de Crédito

Definição: Receitas decorrentes da colocação de títulos públicos ou de empréstimos obtidos junto a entidades estatais ou particulares internas ou externas.

Essas diferenças justificam a criação de dois termos diferentes: **Operações de Crédito** e **Operações Municipais de Crédito**, que, como podemos ver abaixo, além de serem diferentes, pertencem a estruturas temáticas diferentes:

57	Instituições Financeiras
57.01	Bancos
57.01.02	Atividades Bancárias
57.01.02.01	Atividades de Assessoria Financeira
57.01.02.02	Operações Financeiras
57.01.02.03	Operações de Crédito
45	Finanças Públicas Municipais
45.01	Receita Municipal Arrecadada
45.01.01	Receita Municipal Corrente
45.01.02	Receita Municipal de Capital
45.01.02.01	Operações Municipais de Crédito

Através desse exemplo, destacamos a importância do uso das definições na indexação dos dados estatísticos e reafirmamos que a indexação de informação desse gênero de informação não pode se limitar aos passos da indexação listados pela norma NBR 12676, reproduzidos no capítulo 2, parte 2.2.

Além da indexação propriamente dita, no caso de informação estatística, devemos acrescentar a etapa de produção de metadados: definições, notas explicativas, fontes, etc., sendo que para essa produção o documentalista necessita sempre de ajuda dos especialistas na produção e análise de dados estatísticos, por meio de conversas elucidativas e referendo aos textos de definições e notas produzidas.

Pode-se afirmar que o processo de indexação de dados estatísticos é complexo, porque não se restringe às operações de representação como ocorre no tratamento dos textos, mas deve agregar, paralelamente, a identificação de

informações que permitam qualificar e contextualizar cada item de informação. Se na elaboração de uma linguagem específica são utilizados os princípios de elaboração de tesouros, recorrendo-se à terminologia para referendar os termos e suas definições, no processo de indexação são agregadas novas tarefas que se relacionam ao cumprimento das funções contextualizantes e explicativas. Embora constituam processos de natureza diferente, normalmente eles ocorrem de forma paralela, prestando auxílio no processo de tradução conceitual, não sem exigir, posteriormente, readequações na formulação da linguagem.

As definições, notas e fontes são também instrumentos utilizados na etapa de construção da linguagem, mas são especialmente construídas para acompanhar a informação disponibilizada.

Quando indexamos (e quando recuperamos), precisamos de todos os metadados, mas se os termos já existirem no vocabulário, podemos prescindir deles. Na construção da linguagem, eles são importantes para entender os conceitos e agrupá-los semanticamente, mas, uma vez isso feito, a própria estrutura da linguagem acaba por contextualizá-los. Já no processo de análise, os metadados são importantes tanto para a compreensão da informação quanto para a recuperação, permitindo que o usuário, no momento da pesquisa, possa consultá-los para compreender e contextualizar a informação.

A utilização de normas terminológicas permite estabelecer as relações entre os termos da linguagem documentária de forma consistente, pois tais relações se baseiam nas características dos conceitos. A terminologia concreta referencia os termos tais como eles são conceituados nas próprias áreas de especialidade, ao mesmo tempo que oportuniza o fornecimento de definições para os usuários.

Muitas vezes, as necessidades da indexação e as particularidades do sistema de recuperação refletem na construção da linguagem. Para dar exemplo típico, podemos nos referir à indexação dos microdados relativos a *Nascidos Vivos*.

Os microdados, como já foi dito anteriormente, consistem no menor nível de desagregação, retratando na forma de códigos numéricos, o conteúdo dos questionários (Koucher, 2002).

Na Fundação Seade, a base de dados de Nascidos Vivos faz parte de um sistema de tabulações disponível na Internet, que permite aos usuários acessar e construir seus próprios indicadores ou tabelas, referentes às informações do Registro Civil. Os dados dessa base são tirados da Declaração de Nascidos Vivos (Anexo 2).

Quando trabalhamos com microdados, temos de representar informações correspondentes às perguntas do questionário. Como pode ser visto no anexo 2, há uma pergunta sobre o tipo de parto. Ao consultar essa informação na base de microdados, o usuário pode recuperar, para suas tabulações, dados sobre os partos cesáreos, sobre os partos vaginais e sobre os partos cujo tipo é ignorado.

Em relação a esse caso, a primeira decisão, na construção da linguagem, foi a de criar o máximo de termos possíveis para recuperar essas informações: **Parto normal** (que consideramos como um termo mais próximo da linguagem leiga), **Parto vaginal** (cujo conceito abrange os partos com fórceps e partos com vácuo extrator), **Parto cesáreo** e **Cesariana** (como termo equivalente de parto cesáreo). Todos eles, como termos específicos de tipos de Parto.

A estrutura poderia ser apresentada da seguinte forma:

Tipos de Parto – 75.07.01

TT Saúde

TG Parto

TE Parto normal

Parto Vaginal

Parto Cesáreo

UP Cesariana

Na recuperação, entretanto, esse tipo de alternativa se mostrou problemática, pois o sistema de recuperação foi montado de forma a indexar automaticamente todos os termos ascendentes na estrutura da linguagem e mostrar, quando fazemos a busca por qualquer termo, também a estrutura descendente. Isso significa que quando indexamos por **Parto Normal**, automaticamente indexamos por **Tipos de Parto**.

Tendo o sistema de recuperação, essas características, quando a busca era feita por **Tipos de Parto** (termo genérico) ele trazia também os termos da estrutura descendente (termos específicos), tendo a resposta do sistema, a seguinte apresentação:

- TIPOS DE PARTO
☐ Tipo de parto
- PARTO NORMAL
☐ Tipo de parto
- PARTO VAGINAL
☐ Tipo de parto
- PARTO CESÁREO
☐ Tipo de parto

Obs: As palavras que aparecem em maiúsculas são os termos utilizados na linguagem, e as que aparecem depois do quadradinho, são as variáveis a serem selecionadas.

Como se pode ver pelo exemplo, essa visualização mostrou-se inadequada, mas serviu para nos mostrar que na realidade só temos uma variável a ser recuperada, que é o Tipo de Parto. Todos os outros termos referem-se às respostas possíveis de serem recuperadas, não à variável, que é única; portanto, não devem ser utilizados para indexar, evitando uma resposta tão repetitiva e sem sentido, ou melhor, recuperam a mesma informação no formato do microdado.

Como não desapareceu a necessidade de o índice de assuntos permitir recuperação pelos termos mais comuns, como **Parto normal**, **Cesariana**, etc. a solução encontrada foi mudar a estrutura da linguagem, utilizando o recurso do deslocamento genérico, o que alterou a apresentação para:

Tipos de parto – 75.07.01

TT Saúde

TG Partos

UP Parto normal

Parto vaginal

Parto cesáreo

Cesariana

A estrutura assim montada garante que todas as entradas apareçam no índice de assuntos e a recuperação por qualquer uma delas trará, como resultado, a variável que permite extrair do sistema o número de nascimentos com parto normal, ou cesariano ou de tipo ignorado, ocorridos em determinado local e data.

Já quando a busca for feita pelo termo **Tipos de Parto**, somente vai aparecer, como vemos abaixo, o nome da variável a ser selecionada, pois esse termo deixou de ser vinculado a uma estrutura descendente.

TIPOS DE PARTO

☐ Tipo de parto

As dificuldades inerentes à representação de dados estatísticos poderiam ser maiores se não fossem utilizadas as metodologias documentárias e terminológicas.

5 - Considerações Finais

Como pudemos verificar ao longo deste trabalho, a informação estatística obedece a regras particulares. Comparada à informação textual, os dados estatísticos – expressões numéricas que se referem a aspectos da realidade – têm características peculiares que não podem deixar de serem observadas para que se possam constituir sistemas de informação que possibilitem o acesso e a apropriação da informação.

É bastante incomum que esse gênero de informação seja submetido a tratamentos documentários, o que dificulta o trabalho de identificação de procedimentos que permitam sua melhor organização e disseminação, ocorrendo, inúmeras vezes, que as formas utilizadas para sua disponibilização se limitem às normas de apresentação tabular. No entanto, a experiência concreta com o atendimento ao usuário nas buscas por informação estatística evidencia a necessidade de parâmetros para sua organização e tratamento.

As metodologias para o tratamento textual constituíram nosso ponto de partida.

Recuperando a literatura documentária, pudemos observar que os princípios de organização e recuperação da informação têm na noção de linguagem suas bases fundamentais. A informação estatística, embora tenha características distintas da informação textual, também se constitui como linguagem, justificativa principal, portanto, para que nossas investigações tivessem aí sua origem.

As diferenças de tratamento foram vivenciadas ao longo do trabalho e da pesquisa. Dentre tais diferenças, a questão da necessidade da contextualização mostrou-se como um aspecto preponderante. O nível de abstração da informação estatística, enquanto entidade numérica, talvez seja um dos principais indicativos dessa necessidade.

Outras características deste tipo de informação também mereceram reflexões específicas, dentre elas o nível de ordenação das informações. O caso dos microdados mostrou-se revelador, distinguindo-se dentro do conjunto das informações estatísticas analisadas.

Ao final da pesquisa, pudemos sistematizar alguns aspectos básicos, quer relativos às especificidades do dado estatístico, se comparado à informação textual, quer às necessidades de adaptações, tanto no processo de indexação quanto no processo de desenvolvimento de uma linguagem para representá-los.

Com o intuito de elencar os aspectos principais relativos à organização das informações estatísticas, reunimos, no quadro a seguir, as observações que julgamos mais relevantes, desejando que essa sistematização possa auxiliar e orientar àqueles que necessitem dar tratamento documentário à informação estatística.

Quadro 7 – Especificidades da representação da informação estatística

1 – O vocabulário utilizado na representação da informação estatística deve conter termos muito específicos – cada termo praticamente corresponde a um dado indexado.

2 – A linguagem documentária deve ter também a função de organização da informação.

3 – Os conjuntos de informações estatísticas podem ser organizados em grandes temas, seguindo a organização tradicional desse tipo de informação, mas outras hipóteses de organização devem ser construídas, através das estruturas específicas da linguagem e da representação documentária, de forma a reunir informações que se encontram em temas diferentes, mas que juntas formam outros cenários informativos.

4 – É imprescindível construir redes de equivalências e de associações entre os termos, com vistas a adequar a linguagem de organização às solicitações dos usuários, bem como a dirimir suas dúvidas.

5 – É fundamental construir um conjunto de metadados: definições, notas e fontes são indispensáveis, tanto para a contextualização dos dados estatísticos, seja no momento da emissão ou no da recuperação da informação, quanto para assegurar a própria estrutura da organização dos dados.

6 – É preciso ter claro que, mesmo entre os dados estatísticos, poderemos encontrar especificidades que precisam ser levadas em conta na construção da linguagem e no processo de indexação. Essas variações vão surgir de acordo com a forma com que a informação estatística é disponibilizada: tabulações fixas, variáveis para construção de tabulações personalizadas ou microdados para a livre criação de indicadores.

Todavia, além de todos os aspectos citados, a consideração que julgamos mais importante fazer ao final desta pesquisa é a de que a informação estatística, que se caracteriza por suas especificidades em comparação à informação textual, necessita de um tratamento particular que se apresenta como “linguagem especial”, como condição para bem representá-la e recuperá-la. Em síntese: informação específica necessita de linguagem específica.

Referências

AUSTIN, D; DALE, P. **Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngües**. Trad. de Bianca Amaro de Melo e rev. de Ligia Maria Café de Miranda. Brasília: IBCTI; SENAI. 1993. 86p.

BAITELLO JR, N. A sociedade da informação. **São Paulo em Perspectiva**, São Paulo, v.8, n.4, p.19-21, out./dez 1994.

BARRETO, A. de A. A oferta e a demanda da informação: condições técnicas, econômicas e políticas. **Ciência da Informação**, IBICT, v. 8, n.2, 1999. Disponível em: <http://www.ibict.br/>

BARRETO, A. de A. A questão da informação. **São Paulo em Perspectiva**, v.8, n.4, p.3-8, out/dez.1994.

BESSON, J. L. A ilusão das estatísticas. São Paulo, Unesp, 1995. 289p.

CINTRA, A. M. M; TÁLAMO, M. F. G. M; LARA M.L.G; KOBASHI, N. Y. **Para entender as linguagens documentárias**. 2ed. rev. e ampl. São Paulo: Polis, 2002. 92p.

DENCKER, A. F. M; DA VIÁ, S. C. **Pesquisa empírica em ciências humanas: com ênfase em comunicação**. São Paulo: Futura, 2001.190p.

FERREIRA, M. J; TAVARES, I. **Notas sobre a história da Estatística**. Dossiers Didáticos, 4. <http://alea.ine.pt/html/statofic/html/dossier/doc/dossier6.pdf> (consultado em março/2004)

FERREIRA, S. P. Produção e disponibilização de estatísticas. **São Paulo em Perspectiva**, v.17 n.3-4, p.17-25 jul/dez.2003

GRACIOSO, L.S. Disseminação de informações estatísticas no Brasil: prática e políticas das agências estaduais de estatística. **Ciência da Informação**, v.32, n.2, p.69-76, maio /ago.2003.

GUIZZARDI FILHO, O; CONTI, V. L. Produção e disseminação de informações socioeconômicas. **Transinformação**, v.13, n.2, p.43-54, jul/dez. 2001.

IBGE. **A história do Censo no Brasil**. Censo 2000.
www.censo.gov.br/ceiso/censobrasil.shtm (consultado em março 2004).

INGWERSEN, P. **Information retrieval in context**. London: Taylor Graham. 1992.
Disponível em: www.db.dk/pi/iri

ISO 1087-1:2000 (E/F). **Terminology work – Vocabulary, part 1: theory and application/ Travaux terminologiques – Vocabulaire – Partie 1: théorie and application**. 1.ed. Genève: International Standard Organization.

ISO 2788/BS5723 (1986/1987). **Guidelines to establishment and development of monolingual thesauri**. Genève; London: International Standard Organization; British Standard.

JANNUZZI, P. M. **Indicadores sociais no Brasil**. Campinas: Alínea, 2001.141p.

JANNUZZI, P. M; GRACIOSO, L. S. Produção e disseminação de informação estatística: agências estaduais no Brasil. **São Paulo em Perspectiva**, v.16, n.3, p.92-103, jul/set. 2002

JARDIM, Germano G. **A administração pública e a estatística**. Rio de Janeiro: D.I.P., 1941, 169p.

KOUCHER, A. B. **A disponibilização dos resultados do censo de 2002**. Documentos FEE, n.51. Porto Alegre, 2002. Disponível em:
http://www.fee.rs.gov.br/sitefee/download/documentos/documentos_fee_51.pdf

LARA, M. L. G. de. A arquitetura de Sistemas de Informações Estatísticas na Internet. **São Paulo em Perspectiva**, v.12, n.4, p.99-104, out/dez.1998.

LARA, M. L. G. de. Conceitos terminológicos para a construção de linguagens documentárias. (versão preliminar, texto não publicado, 2002).

LARA, M. L. G. de. Linguagem documentária e terminologia. **Transinformação**, v.16, n.3, p.231-240, set/dez 2004

LARA, M. L. G. de. **Representação e Linguagens Documentárias: bases teórico-metodológicas**. 1999. 208f. Tese (doutorado em Ciências da Comunicação). Escola de Comunicações e Artes. Universidade de São Paulo. São Paulo.

LARA, M. L. G; CAMARGO, J. C. C; ROCHA, S. G. Informação estatística e cidadania. **São Paulo em Perspectiva**, v.16, n.3, p.86-91 jul/set. 2002.

LARA, M. L. G; CONTI, V. L. . Disseminação da informação e usuários. **São Paulo em Perspectiva**, v.17 n.3-4, p.26-34, jul/dez.2003

MATOS, M. A. C. **A história da estatística**.
<http://www.amostraestatistica.hpg.ig.com.br/historia.htm> (consultado em março de 2004)

NBR 12676: 1992. **Métodos para análise de documentos – determinação de seus assuntos e seleção de termos de indexação**. Rio de Janeiro, Associação Brasileira de Normas Técnicas - ABNT

PORCARO, R. M. A informação estatística oficial na sociedade da informação; uma (des) construção. **DataGramaZero: Revista de Ciência da Informação**, v.2, n.2, abr. 2001. Disponível em: <http://www.dgz.org.br>.

SAGER, J.C. Prólogo: la terminologia: puente entre varios mundos. In: CABRÉ, M. T. **Terminologia: la teoria, metodologia, aplicaciones**. Barcelona: Antartida/Empariés, 1993.

SENRA, N. de C. Informação estatística: demanda e oferta, uma questão de ordem. **DataGramaZero: Revista de Ciência da Informação**, v.1, n.3, jun.2000.

SENRA, N. de C. Informação estatística: política, regulação, coordenação. **Ciência da Informação**, Brasília, v.28, n.2, 1999. Disponível em: <http://www.ibict.br/>

SENRA, N. de C. Regime e política de informação estatística. **São Paulo em Perspectiva**, v.16, n.3, p.75-85, jul/dez. 2002

SMIT, J. W; BARRETO, A. A. Ciência da informação: base conceitual para a formação do profissional. In: VALENTIN, M. L, org. **Formação do profissional da informação**. São Paulo: Polis, p.9-23,2002.

TÁLAMO, M. F. G. M. **Ciência da informação: informação e conhecimento. Fundamentos da análise documentária. Sistemas de informação, memória e recuperação**. In: Gestão de processos comunicacionais. 2005.

TÁLAMO, M. F. G. M. **Linguagem documentária**. São Paulo: APB, 1997. 12p. (Ensaio APB, n.45).

TÁLAMO, M. F. G. M., et al. Otlet, o criador de estruturas informacionais pela paz mundial. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 20. **Anais (publicação em cd-rom)**, 2002, Fortaleza.

Universidade Federal do Rio Grande do Norte. **História da estatística**
http://www.ccet.ufrn.br/hp_estatistica/historia.html (consultado em março de 2004)

ANEXOS

Anexo 1

TRECHO DA LEI Nº 8.072, DE 25 DE JULHO DE 1990.

Dispõe sobre os crimes hediondos, nos termos do art. 5º, inciso XLIII, da Constituição Federal, e determina outras providências.

O PRESIDENTE DA REPÚBLICA, faço saber que o Congresso Nacional decreta e eu sanciono a seguinte lei:

Art. 1º São considerados hediondos os seguintes crimes, todos tipificados no Decreto-Lei nº 2.848, de 7 de dezembro de 1940 - Código Penal, consumados ou tentados: (Redação dada pela Lei nº 8.930, de 6.9.1994)

I - homicídio (art. 121), quando praticado em atividade típica de grupo de extermínio, ainda que cometido por um só agente, e homicídio qualificado (art. 121, § 2º, I, II, III, IV e V); (Inciso incluído pela Lei nº 8.930, de 6.9.1994)

II - latrocínio (art. 157, § 3º, in fine); (Inciso incluído pela Lei nº 8.930, de 6.9.1994)

III - extorsão qualificada pela morte (art. 158, § 2º); (Inciso incluído pela Lei nº 8.930, de 6.9.1994)

IV - extorsão mediante seqüestro e na forma qualificada (art. 159, caput, e §§ 1º, 2º e 3º); (Inciso incluído pela Lei nº 8.930, de 6.9.1994)

V - estupro (art. 213 e sua combinação com o art. 223, caput e parágrafo único); (Inciso incluído pela Lei nº 8.930, de 6.9.1994)

VI - atentado violento ao pudor (art. 214 e sua combinação com o art. 223, caput e parágrafo único); (Inciso incluído pela Lei nº 8.930, de 6.9.1994)

VII - epidemia com resultado morte (art. 267, § 1º). (Inciso incluído pela Lei nº 8.930, de 6.9.1994)

VII-B - falsificação, corrupção, adulteração ou alteração de produto destinado a fins terapêuticos ou medicinais (art. 273, caput e § 1º, § 1º-A e § 1º-B, com a redação dada pela Lei nº 9.677, de 2 de julho de 1998). (Inciso incluído pela Lei nº 9.695, de 20.8.1998)

Parágrafo único. Considera-se também hediondo o crime de genocídio previsto nos arts. 1º, 2º e 3º da Lei nº 2.889, de 1º de outubro de 1956, tentado ou consumado. (Parágrafo incluído pela Lei nº 8.930, de 6.9.1994)

http://www.planalto.gov.br/ccivil_03/leis/L8072.htm)

Anexo 2

República Federativa do Brasil
Ministério da Saúde
1ª VIA - SECRETARIA DE SAÚDE

Declaração de Nascido Vivo N°

I	Cartório	1	Cartório	Código	2	Registro	3	Data											
		4	Município	5	UF														
II	Local da Ocorrência	6	Local da Ocorrência	7	Estabelecimento	Código													
		1	☐ Hospital	2	☐ Outros estab. Saúde	3	☐ Domicílio												
		4	☐ Outros	9	☐ Ignorado														
		8	Endereço da ocorrência, se fora do estab. ou da resid. da mãe (Rua, praça, avenida, etc)	Número	Complemento	9	CEP												
10	Bairro/Distrito	Código	11	Município de ocorrência	Código	12	UF												
III	Mãe	13	Nome da Mãe	14	Cartão SUS														
		15	Idade (anos)	16	Estado Civil	17	Escolaridade (Em anos de estudo concluídos)	18	Ocupação habitual e ramo de atividade	19	Núm. de filhos tidos em gestações anteriores (obs.: utilizar 99 se ignorado)								
		1	☐ Solteira	2	☐ Casada	1	☐ Nenhuma	2	☐ De 1 a 3	1	☐ Nenhuma	2	☐ De 1 a 3	3	☐ De 4 a 6				
		3	☐ Viúva	4	☐ Separada judicialmente/divorciada	3	☐ De 4 a 7	4	☐ De 8 a 11	4	☐ Cesáreo	9	☐ Ignorado						
		9	☐ Ignorado	5	☐ 12 e mais	9	☐ Ignorado	4	☐ 7 e mais	9	☐ Ignorado								
		20	Logradouro	Número	Complemento	21	CEP												
22	Bairro/Distrito	Código	23	Município	Código	24	UF												
IV	Gestação e Parto	25	Duração da gestação (em semanas)	26	Tipo de gravidez	27	Tipo de parto	28	Número de consultas de pré-natal										
		1	☐ Menos de 22	2	☐ De 22 a 27	1	☐ Única	2	☐ Dupla	1	☐ Vaginal	2	☐ Cesáreo	9	☐ Ignorado				
3	☐ De 28 a 31	4	☐ De 32 a 36	3	☐ Tripla e mais	9	☐ Ignorado	4	☐ 7 e mais	9	☐ Ignorado								
5	☐ De 37 a 41	8	☐ 42 e mais	9	☐ Ignorado	9	☐ Ignorado	9	☐ Ignorado										
V	Recém Nascido	29	Nascimento	30	Sexo	31	Índice de Apgar												
		1	☐ Branco	2	☐ Preta	3	☐ Amarela	4	☐ Parda	5	☐ Indígena	1	☐ M - Masculino	2	☐ F - Feminino	1	☐ 1º minuto	2	☐ 5º minuto
		32	Raça/cor	33	Peso ao nascer	34	Detectada alguma malformação congênita e/ou anomalia cromossômica?												
1	☐ Branco	2	☐ Preta	3	☐ Amarela	4	☐ Parda	5	☐ Indígena	1	☐ Sim	2	☐ Não	9	☐ Ignorado				
35	Polegar direito da mãe	36	Pé direito da criança																
VII	Preench.	37	Responsável pelo preenchimento	38	Função	39	Identidade	40	Órgão Emissor	41	Data								
		Nome																	

ATENÇÃO : ESTE DOCUMENTO NÃO SUBSTITUI A CERTIDÃO DE NASCIMENTO

O Registro de Nascimento é obrigatório por lei.

Para registrar esta criança, o pai ou responsável deverá levar este documento ao cartório de registro civil.