

Bases de dados na economia do conhecimento: a questão da qualidade

Maria de Nazaré Freitas Pereira
Cláudio José S. Ribeiro
Leonel Tractenberg
Priscilla Loureiro Medeiros

“No futuro, as oportunidades irão para os *gatekeepers*, assistentes inteligentes, caçadores de informação especificamente treinados para saber as últimas sobre qualquer assunto.”

(Kerckhove, 1997)

INTRODUÇÃO

Este trabalho decorre de projeto integrado de pesquisa financiado pelo CNPq^{*}, cujo objetivo é produzir conhecimento empírico e metodológico a respeito da qualidade das bases de dados nacionais – um dos pilares da indústria da informação.

Apenas na segunda metade dos anos 90, o tema da qualidade de bases de dados atraiu a atenção da indústria da informação, com trabalhos relacionados à detecção e correção de erros de ortografia e à aplicação de números de autochecagem de dados legíveis por máquina. Essa primeira incursão tinha nas bases de dados bibliográficas seu alvo principal, em decorrência do desenvolvimento de grandes coleções de informação legíveis por máquina, cujo início significativo ocorreu na década de 60^{**}.

A experiência reunida com a produção de bases bibliográficas, as novas tecnologias de informação de base eletrônica e a disponibilidade cada vez mais crescente de outros tipos de bases, como aquelas produzidas pelos agentes de busca da Internet (não-humanos, por excelência), colocam a questão da qualidade como um dos aspectos mais importantes da indústria da informação na atualidade. Uma época marcada pela expansão do público de consumidores de informação, até então circunscrito à esfera de C&T. Os usuários deixam de ser apenas os agentes de C&T e passam a incluir o público em geral, cada vez mais exigente, demandando aquilo que Michael Dertouzos qualifica de “caixinhas preciosas forradas de veludo”[†].

O estado da literatura a respeito do tema da qualidade não permite ainda responder de forma bastante pontual o que deve ser uma base de qualidade. É preciso realizar estudos localizados em componentes dessa indústria, orientados para um uso/conteúdo/contexto em particular. Dois tipos de bases merecem atenção nesse projeto: as bases bibliográficas e as bases tipo diretório/cadastro, que registram a produção científica de pesquisadores brasileiros.

A importância das bases de dados, principalmente das bibliográficas, define-se por seu uso crescente na produção de indicadores de C&T^{††} e, mais recentemente, para produzir estudos estratégicos de C&T, área de conhecimento que se organiza sob a denominação

Resumo

Bases de dados são fundamentais para o processo de análise de C&T em suas relações com o desenvolvimento socioeconômico e com a esfera política mais ampla. Isto implica necessariamente a discussão da questão da qualidade das informações que subsidiam tal análise. Este artigo discute o papel das bases de dados na gestão de C&T para produção de indicadores, mostrando um breve diagnóstico de bases produzidas no Brasil.

Palavras-chave

Bases de Dados; Qualidade; Economia do conhecimento; Gestão de C&T.

^{*} “Por uma economia do conhecimento: avaliação de bases de dados nacionais para a produção de indicadores de C&T (ciência e tecnologia)”.

^{**} O'Neill, Vizine-Goetz (1988).

[†] Dertouzos (1997).

^{††} Observatoire des Sciences et des Techniques (1998).

nação de inteligência competitiva¹. Outro uso é na geração de mapas de conhecimento, que permitem situar graficamente áreas/temas de pesquisa em suas correlações cognitivas, exibindo, ainda, pesquisadores e instituições orientados para a produção desses saberes².

O tema da qualidade comparece neste artigo no bojo de um quadro de referência mais abrangente, onde se destacam: o papel da gestão de C&T na economia do conhecimento, o papel das bases de dados na gestão de C&T, ocasião em que se pontua a qualidade em bases de dados e a experiência francesa na manipulação de grandes bases de dados para produção de indicadores de C&T, um breve diagnóstico das bases de dados brasileiras e o projeto integrado de pesquisa financiado pelo CNPq para fins de estudo da qualidade de um segmento da indústria de bases de dados nacionais.

O PAPEL DA GESTÃO DE C&T NA ECONOMIA DO CONHECIMENTO

Quando da irrupção da doença provocada pela vaca louca na Inglaterra e a contaminação da carne exportada para os países da União Européia, o governo francês solicitou ao Inist (o Ibict francês) um mapeamento do conhecimento produzido a respeito dessa doença para orientar suas decisões. Como o Inist possui uma base de dados bibliográficos com 12 milhões de registros, com controle terminológico e métodos de trabalho conhecidos como infometria, não foi difícil atender à solicitação.

Essa nova versão de funcionamento da sociedade pauta-se no que se convencionou chamar de economia do conhecimento, que, para fins deste trabalho, tem duplo entendimento: de um lado, a definição tradicional que nos diz que esse tipo de economia funciona de forma intensiva com base em conhecimento oriundo de pesquisa científica (Cooper, 1983); de outro lado, nessa economia, o conhecimento sobre o conhecimento, organizado em bases de dados primárias ou de indicadores, é devidamente explorado, fornecendo informações que permitem orientar investimentos governamentais em áreas de C&T, bem como monitorar sua incorporação em setores de produção industrial e de serviços³. Para muitos estudiosos, esse tipo de produção conformaria o que se conhece por sociedade pós-industrial⁴. Para outros, isso define uma sociedade da informação, pois as transações entre seus atores é cada vez mais mediada por produtos e serviços de alto valor informacional, transportados por meios telemáticos, eletrônicos e computacionais⁵.

A relação entre C&T e desenvolvimento começou a ficar evidente quando

estudiosos do campo correlacionaram os efeitos de aumento de investimentos nessas esferas com o crescimento do PIB e dos níveis de emprego, conformando uma nova maneira de produzir riqueza⁶. A correlação tornou-se possível pela disponibilidade de séries de dados estatísticos, tanto do lado da dinâmica de C&T quanto da produção e consumo econômicos.

Essa relação, contudo, com o correr do tempo, mostrou-se não totalmente linear e um novo instrumental teórico e metodológico foi introduzido, definindo, entre outros, um campo de estudos que passou a ser conhecido como indicadores de C&T.

“Indicadores de C&T são conhecimento quantitativo sobre os parâmetros da atividade científica, tecnológica e de inovação aos níveis institucional, disciplinar, setorial, regional, nacional e plurinacional. Tal conhecimento objetiva caracterizar e posicionar instituições, regiões ou países em ‘mapas’ temáticos, permitindo, assim, o estudo comparativo, incluindo análise sobre o tempo” (Barré, 1997b).

No limiar dos anos 70, organizam-se as primeiras experiências no campo: os Estados Unidos, por meio da National Science Foundation (NSF), lançam a primeira edição do relatório “Science and Engineering Indicators”. Na Inglaterra, o “Science Policy Research Unit” (SPRU) e, na França, a divisão de indicadores de um organismo internacional, a OECD, ensaiam os primeiros passos para subsidiar especialistas em C&T com uma série de estudos quantitativos para fins de decisão no âmbito da política de pesquisa de seus países⁷.

Durante as duas décadas seguintes, anos 80 e 90, iniciativas semelhantes

¹ Ver a esse respeito: Dou, H. In which business are we? *Scientometrics*, v.30, p.401-406, 1994. Dou, Henri (1997) *Competitive Technology Intelligence*. In: *Curso de Especialização em Inteligência Competitiva*. Rio de Janeiro: MCT/INT: CNPq/IBICT: UFRJ/ECO. DOU, Henri Jean-Marie. Technology watch and competitive intelligence: the european way. *Competitive Intelligence Review*, v.18, n.1, p.78-84, 1997. Dou, Henry; Dou JR., Jean-Marie. Virtual research univers and research management. *Technology Analysis & Strategic Management*, v.7, n.1, p.109-122, 1995. Faucompré, P., Quoniam, Luc, Dou, Henri (1997) An effective link between science and technology. In: *Sixth International Conference of the International Society for Scientometrics and Informetrics Proceedings*. Israel, Jerusalem. Fuld, Leonard M. (1994) *The New Competitor Intelligence: the complete resource for finding, analyzing, and using information about your competitors*. New York: John Wiley & Sons, Inc. 482 p.

² Muller, Polanco, Royauté, Toussaint (1997).

³ Ver, por exemplo, a introdução e inúmeras análises da principal publicação do Observatoire des Sciences et des Techniques (OST). (1998) e Barré, Laville, Papon (1998).

⁴ Bell (1973).

⁵ Kerckove (1997) e Levinson (1998).

⁶ Machlup (1962) e Porat (1980).

⁷ cf. Barré (1997), op. cit.

em inúmeros outros países europeus e da América Latina testemunham a vitalidade desse campo de estudo e de ação, promovendo mudanças na concepção teórica de indicadores, em suas metodologias de trabalho e, conseqüentemente, em seu formato organizacional, em seus produtos (relatórios de indicadores) e nas decorrentes análises*.

PAPEL DAS BASES DE DADOS NA GESTÃO DE C&T

Um denominador comum a todas essas experiências é o uso intensivo de bases de dados, tanto bibliográficas quanto não bibliográficas. Duas grandes linhas de trabalho, os estudos sociais de ciência e tecnologia e a gestão de C&T, desenvolvem-se construindo análises e embasando argumentos a partir da derivação de indicadores obtidos com o emprego intensivo de dados arrolados nessas bases.

Além disso, o treinamento de pesquisadores e o aumento das relações entre eles continuam tendo nas bases de dados bibliográficas (que se originam do controle da literatura científica) sua principal forma de expressão. Essas bases de dados, ou, na sua ausência, as de natureza cadastral que incorporam referências bibliográficas, permitem conhecer coletivamente o produto intelectual de pesquisadores, bem como a tecedura da rede social em que se sustenta, por meio da construção de indicadores que incorporam outras relações que não apenas as econômicas**.

Verifica-se, portanto, que as bases de dados são fundamentais para a tarefa daqueles que se empenham na análise de C&T em suas relações com o desenvolvimento social e econômico e com a esfera política mais ampla, o que introduz na discussão a questão da qualidade dos dados que subsidiam tais análises.

A questão da qualidade em bases de dados

O controle de qualidade em bases de dados envolve todas as etapas do tratamento da informação, desde a sua criação até o uso final. A qualidade do produto da informação é influenciada pelo *software* de processamento e recuperação, *hardware*, telecomunicações, suporte ao usuário e qualidade do conteúdo da base de dados*. Muitos destes aspectos – dentre eles a engenharia de *software*, a confiabilidade do *hardware* do computador e as telecomunicações – estão fora das abordagens tradicionais da ciência da informação.

Na literatura é possível identificar as duas principais vertentes para aferir qualidade a bases de dados: a qualidade das estruturas de armazenamento e recuperação e a qualidade do conteúdo da base.

Segundo o enfoque da qualidade das estruturas de armazenamento e recuperação, pode-se afirmar que a participação da comunidade de usuários é fundamental na criação de projetos de *software* de alta qualidade. O alcance de tal objetivo demanda a utilização de técnicas contemporâneas de especificação, apoiadas em processos de levantamento e revisão, que permitam agregar qualidade ao projeto.

Na atualidade, as principais técnicas de especificação pressupõem o uso de metamodelos e metateorias**, e essas últimas podem auxiliar a representar, de forma equivalente, os pressupostos das áreas de conhecimento. Os cientistas da informação, em diferentes momentos e contextos, têm adotado diversas formas de representação do conhecimento público em sistemas de informação. Além disto, o uso da técnica JAD*** – Join Application Design – no processo de levantamento e Walkthroughs Estruturados**** no processo de validação são elementos que podem auxiliar o processo de melhoria da qualidade nas estruturas da base de dados.

Além dessas duas técnicas, o uso de quesitos de qualidade* permite medir as características do processo de de-

* O'Neill, Vizine-Goetz (1988).

** Vickery (1997).

*** August (1993).

**** Freedman, Weinberg (1993).

envolvimento, podendo ser incorporados na primeira vertente da qualidade em bases de dados.

Outro aspecto, fundamental para tornar disponíveis bases de dados com qualidade, é o conteúdo que flui nas estruturas de armazenamento e recuperação. Uma base de dados com erros de conteúdo pode trazer grande prejuízo financeiro, e um dado errado pode tornar uma informação incorreta. Os mecanismos de coleta e registro são dimensões de suma importância na qualidade do conteúdo das bases, pois a funcionalidade destas interfaces serão fundamentais no processo de carga da base de dados. As estratégias de crescimento, limpeza dos registros da base de dados e reorganização do conteúdo são fatores que merecem atenção destacada**.

A avaliação do conteúdo dos registros deve ser feita sempre em relação a um conteúdo padrão, definido em consenso pela comunidade de usuários e a partir de fontes de informação confiáveis.

Com base nesta análise, pode-se verificar o atendimento aos objetivos iniciais, ou ainda os requisitos básicos estabelecidos para a base de dados. Por exemplo: pode ser *aceitável* para um grupo de registros de uma determinada base de dados ter um índice de confiabilidade em 90%, mas é *desejável* que este índice seja de 100%. A relação entre o *aceitável* e o *desejável* deve ser tratada na política de qualidade da base de dados, trazendo inclusive as ações de melhoria decorrentes da aferição deste índice.

O estabelecimento de uma política para aferir a qualidade da base, com a coleta de dados sobre o ambiente – uso de recursos, histórico de defeitos e mudanças, reclamações dos clientes –, o uso de dispositivos para verificação e checagem automática de conteúdos contribuirão para a melhoria contínua destas bases. Esta política também deverá contemplar as métricas para avaliação destes conteúdos, permitindo

* Morita-Lou (1984) e Colciencias (1996).

** OECD (1996a) e OECD (1996b).

* Weber (1999).

** Lyon (1976).

do rastrear os melhoramentos nos seus processos e resultados”.

O tópico a seguir nos leva ao encontro de uma das experiências mais interessantes de reunião de bases de dados para produzir conhecimento de outra ordem. A dimensão da qualidade não se coloca de forma explícita, ou seja, não há uma avaliação prévia das bases que serão reunidas para produzir esse conhecimento. Contudo, o trabalho posterior de equivalência dos registros de diferentes bases para produzir compatibilidade entre eles pode ser entendido como o suplemento de força que agrega qualidade às bases de dados.

A experiência francesa: um exemplo de utilização de bases de dados para a produção de indicadores em C&T

Em 1990, ministérios franceses decidem reunir, no Observatoire des Sciences e des Techniques (OST), a função nacional de indicadores de C&T a partir de uma prática organizacional comum entre eles, o GIP (Groupement d'Intérêt Publique). Seu maior objetivo consiste em fornecer amplo diagnóstico do sistema de C&T francês, além de avaliar sua inserção na ciência e tecnologia européia e internacional.

O OST exerce a gestão da informação em ciência e tecnologia por intermédio da exploração sistemática de bases de dados disponíveis no mercado – incluindo bases de patentes, instituições, pesquisadores, nomenclaturas, indicadores, tanto européias quanto norte-americanas – para a operação de uma economia do conhecimento, provedora de informações que permitem orientar investimentos governamentais em áreas de C&T, além de monitorar sua incorporação em setores de produção industrial e de serviços.

Portanto, não é sua função produzir informação básica, mas sim agregar camadas de valor à informação por meio de seu processamento, criando nova informação interna. Esta informação

interna, metodologicamente homogênea, alimenta-se das diferentes fontes para que a comparabilidade entre as informações utilizadas seja garantida, possibilitando análises mediante o relacionamento e cruzamento de dados especializados em uma base de dados central relacional.

O principal produto resultante do trabalho do OST é o *Science & Technologie - Indicateurs*, publicado a cada dois anos. Esta publicação registra pela primeira vez, em 1992, os indicadores da produção científica e tecnológica francesa.

A apresentação do *Rapport des Indicateurs*, bem como seus anexos e alguns trabalhos publicados em periódicos especializados,⁷ permite perceber que a tarefa do OST se apóia em três grandes pilares: bases dados de responsabilidade de outros organismos, um conjunto de *software* para construção de bases de dados relacionais e um enorme trabalho, que demanda edição humana, para tratamento dos dados das bases com as quais opera. Esse trabalho, deveras original, envolve agregação de valor aos dados de cada base por meio de sua reconfiguração e de seu enriquecimento com a introdução de outros dados (como, por exemplo, os códigos postais que permitem reunir a produção de C&T por zonas geográficas), bem como a concepção de indicadores que serão extraídos da base central, indicadores esses sempre em consonância com as políticas de pesquisa, de tecnologia e de inovação.

A experiência de oito anos de trabalho permite a Rémi Barré, diretor do OST, chamar atenção para os limites operacionais concernentes à tarefa de explorar bases de dados, ao mesmo tempo em que propõe “um programa de trabalho para dispor de ferramentas que estejam à altura das ambições de uma economia do conhecimento”, em outras palavras, atenção para a dimensão da qualidade, cujo objetivo é:

“(…) proceder a toda uma série de experimentações sobre as bases de dados em questão para adquirir um bom conhecimento de seu conteúdo para estabelecer, precisamente, as condi-

ções de validade de sua utilização”.

Nesse sentido, sua proposta merece ser considerada por aqueles que de uma forma ou de outra se envolvem com indicadores. Seu programa de trabalho destaca os seguintes itens:

a) Ampliar a oferta de bases de dados primárias para evitar monopólio, pois as bases que suportam os indicadores são vendidas. Outro aspecto diz respeito a critérios pouco transparentes na inclusão de dados nas bases e, nesse sentido, deveria haver um estímulo à produção de bases que atendessem às demandas dos produtores de indicadores.

b) Enriquecer as bases de dados existentes a partir de um trabalho de compatibilidade de nomes (instituições), identificação de afiliações institucionais, correspondência entre o endereço do autor e códigos geográficos da região em que se situa sua instituição de trabalho.

c) Estabelecer relações entre as bases de dados de patentes e as bases de dados institucionais.

d) Efetuar um intenso trabalho metodológico envolvendo teste, elaboração, difusão e validação das bases, sua dimensão qualitativa.

e) Inovar no plano institucional para construir coletivamente o “grande equipamento bibliométrico”, que requer negociação a respeito do uso e acesso às bases de dados e melhor cooperação entre seus produtores, assim como descentralização na produção das bases de dados, cabendo sua construção à instituição que detém o domínio da matéria a ser coberta pela base.

O programa de trabalho proposto por Rémi Barré se constitui na principal inspiração do projeto do qual decorre este artigo e nos remete ao próximo tópico no qual se indaga a respeito das bases de dados nacionais para produção de indicadores em C&T.

BREVE DIAGNÓSTICO DAS BASES DE DADOS BRASILEIRAS

A análise da “realidade” das bases nacionais aqui apresentada baseia-se em

⁷ Arthur (1994).

⁷ Barré (1997a).

uma tipologia provisória para a extração de dados de um cadastro de bases de dados produzidas por instituições públicas e privadas brasileiras, publicado pelo IBICT. O registro de informação sobre cada base contém os seguintes campos: nome da base; sigla; tipo de informação; tipo de acesso; meios de consulta; assunto; área(s) geográfica(s).

A dimensão escolhida para este diagnóstico inicial fornece pistas da situação de bases de dados factíveis de serem utilizadas na produção de indicadores de C&T.

O cadastro arrola grande quantidade de bases de dados produzidas pelo IBGE e pela Cetesb (São Paulo). Decidiu-se deixar o conjunto de bases de dados de ambas sem classificar, ainda que algumas pudessem ser enquadradas com uma extensão da tipologia proposta. Contudo, como o registro das bases de ambas as organizações não nos indica claramente se as bases que produzem são do tipo bibliográfica ou cadastral, decidiu-se classificar os dois conjuntos na categoria OUTRAS BASES, subdividindo-se em IBGE, Cetesb – para expressar a produção integral de ambas.

A tabela 1 apresenta um total de 694 bases de dados produzidas por 245 instituições representadas no referido cadastro. Na tabela, as bases de dados são caracterizadas sob o ponto de vista do conteúdo/produto que veiculam.

A distribuição do total de 694 bases ocorre conforme os seguintes tipos: bases de produção técnico-científica, cadastros e guias, bases de acervo, bases de informação tecnológica, de negócios e de gestão tecnológica, bases de culturas vivas, bases de indicadores/estatísticas/diagnósticos e outras bases.

Em cada uma dessas categorias, o levantamento apresenta, em ordem decrescente, o seguinte resultado: bases de acervo – 272 (39%); cadastros e guias – 142 (20%); bases de produ-

TABELA 1
Bases de dados produzidas no Brasil

Classificação	No. de BD's
A - Base de Produção Técnico-científica	
Bibliográficas	077
Teses	012
Subtotal	089
B - Cadastros e Guias	
Gerais	035
Instituições	022
Empresas	030
Pesquisadores	016
Produtos, serviços e espécies	023
Eventos	006
Filmes e vídeos	006
Outros	004
Subtotal	142
C - Bases de Acervo	
Gerais	234
Imagens	004
Legislação	019
Recortes/notícias	003
Autoridades/terminológica	012
Sub-total	272
D - Bases de Informação Tecnológica, de Negócios e de Gestão Tecnológica	
Gerais	020
Normas técnicas	011
Patentes	004
Subtotal	035
E - Bases de Culturas Vivas	
	008
F - Bases de Indicadores/Estatísticas/Diagnósticos	
	012
G - Outras Bases	
Gerais	10
IBGE	78
Cetesb	48
Subtotal	136
TOTAL	694

Fonte: Instituto Brasileiro de Informação em Ciência e Tecnologia. (1996), *Bases de Dados Produzidas no Brasil*. Brasília. 95p. Documento elaborado pelo Departamento de Disseminação da Informação (DDI) para a Reunião da Comissão Temática Sistema de Informação Científica Tecnológica do Mercosul, Buenos Aires - Argentina, 15 a 16 de abril de 1996.

ção técnico-científica – 89 (13%); bases de informação tecnológica, de negócios e de gestão tecnológica – 35 (5%); bases de indicadores/estatísticas/diagnósticos – 12 (2%); bases de culturas vivas – 8 (1%); e outras bases – 36 (20%).

As bases de acervo constituem a maioria (39% do total) e subdividem-se em gerais, imagens, legislação, recortes/notícias, autoridade/terminologia. São bases que refletem, por excelência, os catálogos das bibliotecas, elo indispensável entre seus usuários e os acervos de livros, títulos de periódicos, mapas, leis, recortes etc. Merece registro a presença nesta categoria de 12 bases de autoridade/terminologia, ferramentas de trabalho do bibliotecário/especialista de informação, fundamentais

para a tarefa de descrever acervos de forma padronizada: as bases de autoridade. Estas bases, conhecidas também como “nome certo”, permitem registrar, de forma consistente, o nome de um autor ou instituição, estabelecendo relações entre eles ou variações em sua forma. As bases de terminologia, por sua vez, apresentam os termos usados para descrever os assuntos do acervo, bem como as relações entre eles, refletindo áreas de conhecimento com suas hierarquias e interdependências.

Merece registro o fato de as bases de autoria e de terminologia representarem uma importante contribuição, principalmente como *know-how*, para o

estabelecimento do que chamamos de suplemento de força à tarefa de agregar bases de dados primários em uma grande base de dados relacional¹.

O segundo conjunto de bases de dados mais expressivo numericamente é do tipo cadastros e guias com 20% do total e cuja função principal é informar “quem é quem” ou o “que é o quê” em determinado setor de atividade humana. Elas subdividem-se em gerais, instituições, empresas, pesquisadores, produtos/serviços/espécies, eventos, filmes e vídeos e outras.

Vejamos, agora, cada uma das subdivisões do conjunto cadastros e guias em separado.

As bases de dados gerais são aquelas conhecidas como bases de fontes de informação em determinada área de interesse. Elas passaram a ser produzidas com o apoio de uma metodologia produzida pelo IBICT/DCD, o que, de certa forma, ajuda a entender sua presença significativa (35 bases em um total de 142) na categoria cadastros/guias. Estas bases arrolam informações sobre instituições, pesquisadores, laboratórios, normas técnicas, dicionários e outras fontes, formando um híbrido cadastro-bibliografia.

As bases de dados de instituições (22 ocorrências) arrolam informações sobre instituições públicas/organismos governamentais, incluindo, ainda, bases de dados sobre sociedades científicas e laboratórios governamentais. Merece registro a ocorrência de uma única base de sindicatos.

A subcategoria seguinte, empresas (30 registros), informa sobre empresas estatais e privadas do setor industrial, comercial e de serviços.

Os cadastros/guias de pesquisadores compõem com um total de 16 registros. Basicamente, constam de informações sobre os recursos humanos

das instituições produtoras da base de dados ou da área de conhecimento de seu interesse, como é o caso da *publicação Quem é Quem em Botânica*, editada pela Sociedade Brasileira de Botânica.

Em seguida, tem-se uma subdivisão deveras interessante, qual seja, a de produtos, serviços e espécies. Essas bases registram produtos desenvolvidos por empresas, bem como serviços por elas prestados, geralmente empresas industriais e laboratórios. Faz-se representar aqui bases de dados que discriminam espécies animais, minerais e florestais, bem como aquelas que descrevem localidades em que ocorrem riquezas minerais. Algumas bases de dados do IBGE e da Cetesb, mantidas em seu conjunto em outras bases, são desse tipo. Pode parecer estranha a presença de tais bases no IBGE, porém é bom lembrar que este Instituto já desenvolveu funções de geografia (meio ambiente). Pelo mesmo motivo, estas bases ocorrem também na Cetesb.

As bases de dados de eventos (seis registros), por sua vez, veiculam informações sobre congressos, seminários, conferências e reuniões especializadas realizadas no Brasil e/ou no exterior.

A categoria seguinte de bases do tipo cadastros/diretórios é a de filmes e vídeos com seis registros. Trata-se de bases que referenciam e descrevem títulos de filmes e vídeos sobre um dado assunto/tema/área de atuação, não veiculando imagens. As bases de imagens são bases de acervo e aparecem na categoria correspondente.

Finalmente, a última ocorrência da categoria cadastros/guias registra quatro bases como “OUTRAS”. Aqui estão reunidas bases de linguagens documentárias (descrevem as ferramentas – o tesouro, por exemplo – utilizadas pelo bibliotecário para indexar o acervo, e podem indicar quem as utiliza) e bases de dados que arrolam projetos de uma instituição. Algumas vezes, essas bases juntam projetos à produção técnico-científica dos pesquisadores/especialistas da instituição produtora da base de dados. Nesse caso, a base de dados está classificada na

categoria base de produção técnico-científica. Como exemplo, tem-se a Base de Projetos da Codevasf e a Base Paleontologia da CPRM, que registra cada projeto e a produção dele decorrente.

A próxima categoria em ordem quantitativa está representada pelas bases de produção técnico-científica que compõem nesse levantamento com 89 ocorrências, sendo 77 bases de produção técnico-científica propriamente dita e 12 bases de teses e dissertações. Essas bases são de grande interesse neste levantamento, pois são exatamente elas que nos permitem começar a indagar a respeito da questão: a quantas anda o registro intelectual da produção bibliográfica científica nacional em bases de dados? Ele é significativo? As bibliografias nacionais ainda são produzidas? Como se organiza o registro dessa produção? As bases de dados bibliográficas tradicionais, se existentes, carecem de qualidade de cobertura/descrição, podendo ser desprezadas? As bases de controle da produção científica, sejam aquelas do tipo bibliografia ou do tipo cadastro – CNPq (diretório de pesquisa), da Capes (controle da pós-graduação) e a recentemente organizada pelo PADCT –, substituem essas bibliografias tradicionais? A tecnologia da informação pode ajudar na tarefa de reuni-las? Isso faz sentido?

Vejamos a primeira questão: a quantas anda o registro intelectual da produção bibliográfica científica nacional em bases de dados?

As 77 bases de dados incluídas na categoria Produção Técnico-Científica apresentam níveis de organização bastante diferenciados. Alguns exemplos nos ajudam a entender essa particularidade:

- relatórios técnicos de um grupo de pesquisa (pesquisadores do Cetem), ou de uma área de conhecimento (área nuclear coberta pela base do CIN/Cnen);
- artigos de periódicos nacionais e estrangeiros em áreas definidas de acordo com interesses de grupos de pesquisa (tecnologia mineral, química,

¹ Faz parte da proposta da coordenação de indicadores do MCT/CNPq a consolidação das bases de dados de indicadores de C&T do CNP em um banco de dados Oracle (CNPq, 1997).

engenharia química, tratamento de minérios, meio ambiente; todas essas áreas de interesse de grupos do Cetem);

- artigos de periódicos que são editados pela instituição de pesquisa (é o caso das revistas científicas *Acta Amazonica* e *Amazoniana* publicadas pelo Inpa);
- artigos de periódicos e trabalhos publicados em anais de seminários/congressos e reuniões especializadas (é o caso da base PERI, produzida pela Escola de Biblioteconomia da UFMG, cobrindo área de ciência da informação, biblioteconomia, administração e computação);
- literatura abrangendo áreas geográficas e de assuntos delimitados (América Latina/Caribe e as ciências da saúde, base da Bireme; Brasil e a produção da odontologia, base da Faculdade de Odontologia da USP);
- literatura abrangendo instituições ligadas à pesquisa e assuntos delimitados (produção acadêmica das inúmeras faculdades da USP).

É fato que as bases do tipo bibliográficas tradicionais e do tipo produção científica podem ser consideradas bases bibliográficas. Contudo seus métodos de produção diferem: as bases de produção científica têm origem no controle das publicações científicas de grupos de pesquisa, apresentando forte orientação institucional. Por sua vez, as bases bibliográficas têm origem no controle da literatura científica, principalmente a periódica, apresentando forte orientação temática.

Há evidências de que no Brasil proliferaram as bases de controle da produção científica. Supomos que isso ocorre para facilitar o trabalho anual de preenchimento da base Capes – base que arrola a produção científica dos programas de pós-graduação no país.

As bibliografias tradicionais têm sua origem na criação do IBBD. Seu serviço de bibliografia se encarregava de registrar a produção científica arrolada

nos periódicos nacionais, produzindo bibliografias de grandes áreas de conhecimento, como física, matemática, química, entre outras, com intenso trabalho de indexação. O método de trabalho que gerava essas bibliografias permitia produzir, ainda, o cadastro de pesquisadores e as pesquisas em processo. Com a criação do IBICT, essas bibliografias foram descontinuadas, e, com o correr do tempo, ele passou a produzir metodologias para guias de fontes de informação, incluindo, em uma só fonte, bibliografias (não necessariamente produzidas no país), cadastros e outras fontes.

O controle da literatura científica (como bibliografia e não como cadastro) é uma atividade bem-sucedida apenas em três grandes áreas de conhecimento: agricultura (Base Agrícola), energia nuclear (Base INIS/CIN) e ciências da saúde (Base LILACS/Bireme). Pelo menos isso é que é fato e bem sabido.

O conhecimento produzido no Brasil não se restringe às áreas anteriormente citadas. E as bases tipo diretório/cadastro não podem ser consideradas como substitutas das bases bibliográficas.

É fato que os indicadores de produção científica no Brasil têm sua origem nas bases cadastrais⁷. Isso gera como resultado tabelas e análises por demais gerais (ciências exatas e da terra, ciências biológicas, engenharias etc.), o que impossibilita qualquer decisão política orientada para temas/questões de conhecimento mais localizados.

Não é possível fazer mapas de conhecimento e nem inteligência competitiva a partir de bases de dados cadastrais já que estas não são orientadas por áreas/temas/questões/métodos/controvérsias de conhecimento bastante pontuais. Sua função pode ser, quando muito, complementar. Por outro lado, parece fora de propósito explorar apenas as bases de dados internacionais para tomar decisões estratégicas locais.

A relativa escassez, parcialidade, generalidade e pouca abrangência de informações sobre a produção de C&T no

país pode ser atribuída, entre outros fatores, à deficiente cobertura da literatura brasileira (ausência de um controle bibliográfico nacional) – com exceção das áreas de ciências da vida, energia nuclear e agricultura – e à proliferação de bases departamentais de produção científica, gerando comprometimento da qualidade dos dados, incompatibilidades interbases, assim como baixa visibilidade da produção acadêmica.

Como conseqüências, são geradas análises e avaliações distantes da realidade, acarretando, portanto, resultados pouco construtivos para o aperfeiçoamento do sistema de C&T do país.

O diagnóstico aqui apresentado foi incluído no projeto submetido ao CNPq, funcionando para justificar o projeto de pesquisa aprovado, ora em seu primeiro ano de execução.

O PROJETO BRASILEIRO PARA A CONSTRUÇÃO DE UMA ECONOMIA DO CONHECIMENTO

No sentido de buscar soluções para esta escassez de informações sobre a produção de C&T no país, devido à deficiente cobertura da literatura brasileira e ausência de um controle bibliográfico nacional, torna-se necessário pensar em uma agenda que, aproveitando-se da experiência francesa, inclua propostas de ação adaptadas à nossa realidade, mais especificamente, que propicie:

1) conhecimento das implicações para os estudos de política científica da falta de utilização de bases de dados bibliográficas tradicionais (que se originam do controle da literatura) para a extração de indicadores de C&T;

2) levantamento dos problemas decorrentes do registro crescente da produção científica nacional em bases cadastrais, que se estruturam a partir da demanda de financiamento (CNPq (BCUR) e PADCT) e de avaliação (Capes), e em bases de produção científicas locais, que se estruturam a partir do controle interno de seus produtores;

⁷ Brasil. Ministério da Ciência e Tecnologia (1996); Brisolla (1998); Ricyt (1997).

⁸ CNPq (1993).

3) avaliação sobre em que medida as bases cadastrais e de produção científica departamental podem se constituir em substitutivo de bases de dados bibliográficas que se estruturam a partir do controle da literatura, propiciando a constituição da memória científica nacional.

Neste intuito, o projeto integrado de pesquisa “Por uma economia do conhecimento: avaliação de bases de dados nacionais para a produção de indicadores de C&T (ciência e tecnologia)” representa uma iniciativa atual cujo objetivo principal é estudar a questão da qualidade em bases de dados nacionais e sua influência na construção da economia do conhecimento.

CONSIDERAÇÕES FINAIS

A importância das bases de dados na economia do conhecimento poderia ser suficiente para convencer aliados da importância de realizar o projeto aqui proposto. Contudo, ele se sustenta, também, em outras duas ordens de razão: uma política, qual seja a de atender recomendação do Conselho de Ciência e Tecnologia (CCT) da Presidência da República quanto à avaliação das bases de dados nacionais que sustentam a produção de indicadores de C&T; a outra razão é teórica – a da relação entre C&T e desenvolvimento econômico e o papel que os indicadores de C&T e suas bases de dados desempenham na definição daquilo que se denomina economia do conhecimento, economia da informação, sociedade da informação ou sociedade pós-industrial. Esta razão, que tem seu pilar na economia, não reina absoluta e, longe de relegar a importância das bases de dados para segundo plano, contribui ainda mais para valorizá-las, principalmente em sua dimensão qualitativa.

REFERÊNCIAS BIBLIOGRÁFICAS

1. ARTHUR, Lowell J. *Melhorando a Qualidade do Software: Um guia para o TQM*. Tradução de Flávio Eduardo Frony Morgado. Rio de Janeiro: Infobook, 1994. p.144.
2. AUGUST, Judy H. *JAD: Join Application Design*. Tradução Marcelo Melo Molinari. Revisão técnica de Ronaldo Stevis Cassiolato. São Paulo: Makron Books, 1993.
3. BARRÉ, Rémi. Les Outils bibliométriques, instruments essentiels pour les recherches concernant les économies fondées sur la connaissance. *Revue d'Économie Industrielle*, n.79, 1997a. p.119-128
4. BARRÉ, Rémi. The European perspective on S&T indicators. *Scientometrics*, v.38, n.1, 1997b. p.57-70. p.58.
5. BARRÉ, Rémi; LAVILLE, Françoise; PAPON, Pierre. *Quelques indicateurs de la compétitivité technologique de la France*. Paris: Observatoire des Sciences et des Techniques, 1998.
6. BELL, Daniel. *O advento da sociedade pós-industrial*. São Paulo: Cultrix, 1973. 540p.
7. BRASIL. Ministério da Ciência e Tecnologia. *Indicadores nacionais de Ciência e Tecnologia: 1990-1994*. Brasília: MCT, 1996. 45p.
8. BRISOLLA, Sandra. *Ciência e Tecnologia no Estado de São Paulo*. Campinas: UNICAMP - Departamento de Política de Científica, 1998. (Estudo encomendado pela FAPESP).
9. COLCIENCIAS. (1996). Hacia la construcción de un Observatorio de Ciencia y Tecnología. Bogotá. 159p.
10. CNPq. *O fomento do CNPq nos estados e instituições de pesquisa*. Brasília: MCT/CNPq, 1993.
11. CNPq. *Diretório dos grupos de pesquisa no Brasil*. v.1. Brasília: MCT/CNPq, 1995.
12. CNPq. *Termo de referência preliminar: consolidação das bases de dados de indicadores de C&T do CNPq*. Brasília: MCT/CNPq, Coordenação de Indicadores de C&T. 1997.
13. COOPER, Michel. The structure and futur of the information economy. *Information Processing & Management*, v.19, n.1, 1983. p.9-26.
14. DERTOUZOS, Michael. *O que será; como o novo mundo da informação transformará nossas vidas*. São Paulo: Companhia das Letras, 1997.
15. INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. *Bases de Dados Produzidas no Brasil*. Brasília: IBICT, 1996. 95p.
16. FREEDMAN, Daniel P.; WEINBERG, Gerald M. *Manual de Walkthroughs: inspeções e revisões técnicas de sistemas e programas*. Tradução Reinaldo Castello, revisão técnica Ronaldo Stevis Cassiolato. São Paulo: Makron Books, 1993.
17. GONZALEZ DE GOMEZ, Maria Nelida. O papel do conhecimento e da informação nas informações políticas ocidentais. *Ciência da Informação*. v. 16, n. 2, jul./dez., 1987. p. 157-167
18. KERCKHOVE, Derrick de. *A pele da cultura*. Lisboa: Relógio D'Água, 1997. 294p.
19. LEVINSON, Paul. *A arma suave: história natural e futuro da revolução da informação*. Tradução de J. Freitas e Silva. Lisboa: Editorial Bizâncio, 1998. 287p.
20. LYON, John K. *The Database Administrator*. New York: John Wiley & Sons, 1976 p. 86-99
21. MACHLUP, Fritz. *The production and distribution of knowledge in the United States*. New Jersey: Princeton University Press, 1962.
22. MORITA-LOU, Hiroko (ed.). Science and technology indicators for development. Proceedings of the Panel of Specialists of the United Nations Advisory Committee on Science and Technology for Development. Graz, Austria, 2-7, May 1984. London: Westview Press, 1984. 207p.
23. MULLER, Chantal; POLANCO, Xavier; ROYAUTÉ, Jean; TOUSSAINT, Yannick. *Acquisition et structuration des connaissances en corpus: éléments méthodologiques*. Lorraine: INRIA, 1997. Rapport de Recherche, 3198
24. O'NEILL, Edward T., VIZINE-GOETZ, Diane. Quality Control in online Databases. *ARIST*, New Jersey, v.23, 1988. p.125-156.
25. OECD. New indicators for the knowledge-based economy: proposals for future work. In: CONFERENCE ON NEW S&T INDICATORS FOR THE KNOWLEDGE-BASED ECONOMY, 1996. 17p. (DSTI/STP/NESTI/GSS/TIP (96) 6).
26. OECD. Summary record of the Conference held on 19-21 June 1996. In: CONFERENCE ON NEW S&T INDICATORS FOR THE KNOWLEDGE-BASED ECONOMY, 1996. 11p. (DSTI/STP/NESTI/GSS/TIP (96) 5).
27. OBSERVATOIRE DES SCIENCES ET DES TECHNIQUES. *Science & technologie; indicateurs 1998*. Paris: Economica, 1998. 551p.
28. PORAT, Marc Uri. *The information economy: definition and measurement*. Washington: United States Department of Commerce, Office of Telecommunications, 1980.
29. RICYT. *Indicadores de Ciencia y Tecnología Iberoamericanos/Interamericanos: 1990-1996*. Buenos Ayres, 1997. 87p
30. VICKERY, Brian. “Metatheory and Information Science”. *The Journal of documentation*, v.53, n.5, 1997. p.457.
31. WEBER, Kival C. e ROCHA, Ana R. C. *Qualidade e Produtividade em Software*. São Paulo: Makron Books, 1999.

Databases in the knowledge economy: the issue of quality

Abstract

Databases play a major role on the analysis of Science & Technology development and its relations to socioeconomic and political fields. This fact brings up the discussion on the quality of information being used. The present paper discusses the role of databases in the production of S&T indicators, mentioning the French experience concerning S&T analysis and gives a preliminary overview of the databases being produced in Brazil.

Keywords

Databases; Quality; Knowledge economy; Science and Technology Management.

Maria de Nazaré Freitas Pereira

Doutora em ciências humanas (IUPERJ). Pesquisadora titular do Departamento de Ensino e Pesquisa do IBICT, com pós-graduação em ciência da informação.

E-mail: marianazare@uol.com.br

Cláudio José S. Ribeiro

Engenheiro, analista de sistemas e mestrando em ciência da informação. ECO/IBICT/UFRJ.

Leonel Tractenberg

Psicólogo e matemático pela UFRJ. Bolsista de apoio técnico (CNPq).

Priscilla Loureiro Medeiros

Estudante de biblioteconomia (ênfase em gestão da informação) pela Universidade Santa Úrsula. Bolsista de Iniciação Científica (CNPq).
