

Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T*

Carlos Henrique Marcondes

Doutor em Ciência da Informação.
Universidade Federal Fluminense, Departamento de Ciência da Informação.
marcondes@alternex.com.br

Luis Fernando Sayão

Doutor em Ciência da Informação.
Comissão Nacional de Energia Nuclear/Centro de Informações Nucleares.
lsayao@cnen.gov.br

Resumo

São discutidos os novos mecanismos de cooperação entre sistemas de informação em ciência e tecnologia surgidos a partir da emergência de arquivos de publicações digitais livres – os **open archives** – e de mecanismos de interoperabilidade entre estes arquivos, como o protocolo **Open Archives Initiative Protocol for Metadata Harvesting**. É apresentado um histórico do surgimento das publicações digitais e dos arquivos abertos como uma alternativa aos mecanismos tradicionais de comunicação científica, baseados nos periódicos acadêmicos tradicionais. São discutidas as novas formas de cooperação e os novos serviços informacionais viabilizados a partir do surgimento dos **open archives** e das tecnologias a eles associados. As consequências e oportunidades desta situação para a ciência e para os sistemas e redes de informação dos países em desenvolvimento são analisadas.

Palavras-chave

Comunicação científica; Bibliotecas digitais; Documentos digitais; Informação em ciência e tecnologia; Cooperação e compartilhamento; Arquivos abertos; Interoperabilidade; Protocolos

Digital documents and new ways of cooperation among information systems

Abstract

This work discusses new mechanisms of cooperation among Science, Technology and Medicine information systems arisen with the emergence of electronic archives of free digital publications – the open archives – and of interoperability standards between them as the Open Archives Initiative Protocol for Metadata Harvesting. A historical scheme of the emergence of electronic publishing and open archives as alternative mechanism of scientific communication to the traditional academic journals is presented. New information services and cooperation schemas brought about by the raising of open archives are discussed too. Finally, consequences and opportunities of this situation to developing countries are analyzed.

Keywords

Scientific communication; Digital libraries; Digital documents; Science, technology and medicine information; Cooperation and resource sharing; Open archives; Interoperability; Protocols.

INTRODUÇÃO

É inquestionável o papel central que desempenham hoje as tecnologias de informática, computação e comunicação nas práticas de informação (Marcondes, 1999). Quando se fala em informação para ciência e tecnologia, este papel é mais acentuado ainda. Isto porque a ciência institucionalizada está assentada em mecanismos de comunicação rápida dos resultados de pesquisa, que por sua vez estão hoje baseados fortemente nas tecnologias de informação. No ciclo de comunicação científica, as bibliotecas têm um papel fundamental. A elas cabem, neste ciclo, os papéis de coleta, registro, estocagem e disseminação de informações. A evolução das tecnologias de informação, no entanto, vem alterando substancialmente este papel e junto com isto o próprio conceito de biblioteca.

Um aspecto problemático da cultura de nosso tempo relacionado à questão informacional é o assim chamado fenômeno da explosão informacional, a grande quantidade de informações produzidas e disponibilizadas por diferentes atividades sociais, dificultando sua identificação, acesso e utilização. Desde a emergência deste fenômeno, a partir da metade do século XX, a saída encontrada pelas bibliotecas foi a cooperação: a publicação das fichas do catálogo da Library of Congress (EUA) data de 1905, o projeto MARC é do fim da década de 60 (Robredo, 1994). Desde a invenção do computador na década de 50, as *tecnologias de informação passaram a ser usadas pelas bibliotecas para prover acesso não só a documentos dos seus próprios acervos, mas também aos armazenados em acervos de outras bibliotecas* (Marcondes, 2001a). Catálogos coletivos e bases de dados automatizadas, mecanismos de empréstimo e provisão de cópias digitalizadas de documentos entre bibliotecas são exemplos do uso de tecnologias de informação para prover acesso a documentos (ainda em papel), disponíveis em acervos para além dos muros de cada biblioteca.

* Trabalho apresentado na Reunião de Especialistas em Informação Científica Digital, promovida pela Bireme/OPS/OMS e Unesco, São Paulo, 26-27 de março de 2002.

O surgimento da Internet a partir dos anos 90 vem mudando de maneira radical o papel das bibliotecas no ciclo intermediação e acesso ao documento. As possibilidades abertas pela Internet com seus mecanismos de publicação direta na rede tornam o acesso a um documento digital uma mera questão de conhecer sua URL. No entanto, esta facilidade de acesso tem como contrapartida a grande dificuldade de encontrar informação relevante, as atividades de *information discovery*. Encontrar a informação relevante é fundamental para que a mesma possa ser utilizada. O uso dos mecanismos de busca gerais é uma solução parcial para o problema. Sneiderman (1997), discutindo a eficácia dos mecanismos de busca de uso geral disponíveis na Internet, diz que:

*“Embora mecanismos de busca como as Infoseek, AltaVista, Lycos, WebCrawler e Open Text sejam largamente usados, existe um consenso público e geral e entre profissionais acerca das grandes dificuldades em se buscar informações”**

Um relatório do NEC Research Institute publicado pelo boletim *Edupage*, em português, datado de 3 de abril de 1998, relata que, na ocasião, os melhores mecanismos de busca cobriam não mais de 30% de todas as páginas Web. Esta afirmação reforça a dimensão do problema localização/identificação colocado pela Internet, devido principalmente ao seguinte fator: O estudo mais recente de Bergman (2001) sobre a “Deep Web”X “Surface Web” enfatiza a gravidade da questão.

- grande quantidade de informações e seu número cresce de forma acelerada; acentua-se a chamada “explosão informacional”, existem bilhões de páginas; publica-se de tudo, sobre tudo, de forma caótica, na Internet;
- a Internet não é como uma biblioteca: não existe ordem, a informação é disponibilizada de maneira caótica;
- a informação disponibilizada é sobre uma infinidade de temas, sob os mais diferentes enfoques; isto é, um problema adicional para sua recuperação; não é à toa que os profissionais de informação, principalmente aqueles que trabalham com informação especializada para ciência e tecnologia, desenvolveram técnicas como vocabulários controlados e linguagens artificiais, técnicas de indexação pós-coordenada, conectivos booleanos para coordenar conceitos temáticos simples, formando conceitos temáticos complexos etc.;

– a informação é disponibilizada em diferentes idiomas, o que agrava o problema do controle do vocabulário a ser usado na recuperação;

– os mecanismos de busca gerais indexam a Internet periodicamente, de forma automática, cegamente, sem compreender o tema de uma página, simplesmente extraindo palavras do texto HTML da página e armazenando estas palavras isoladas junto ao endereço da página, em uma base de dados para consulta. Além disso, a indexação é feita por páginas HTML isoladas, não considerando que diversas páginas estão inter-relacionadas, formando um *site*;

– os programas-robôs dos mecanismos de busca só “enxergam” páginas HTML estáticas quando fazem sua rotina de indexação, deixando de considerar grande quantidade de informações sob a forma de registros contidos em bases de dados disponíveis na Internet – a chamada *deep web* – (Bergman, 2001), que fica assim “invisível”. Estes registros são acessados somente por meio das interfaces destas bases de dados, que pressupõem uma interação entre um usuário humano e a base de dados e, portanto, ficam inacessíveis aos programas-robôs. Segundo Bergman, a *surface web* é estimada em cerca de 2,5 bilhões de páginas, enquanto a *deep web* seria cerca de 500 vezes maior;

– tem sido extremamente difícil estabelecer filtros de qualidade para as informações encontradas na Internet.

O surgimento da Internet vem tendo um impacto significativo também nas formas de comunicação científica e, conseqüentemente, nos sistemas de informação em C&T. Diferentes processos sociais, econômicos e tecnológicos convergem para configurar a situação atual das formas de comunicação científica. Desde o surgimento do primeiro arquivo eletrônico de *preprints*, ou *eprints*, o ArXiv, no Los Alamos National Laboratory, criado em 1991 pelo físico Paul Ginsparg (Ginsparg, 1996), que a própria comunidade científica internacional oferece uma alternativa prática para a publicação de seus trabalhos e sua disponibilização gratuita como alternativa aos periódicos científicos controlados pelos grandes editores internacionais. Este debate envolvendo pesquisadores e grande editores vem sendo coberto pela revista *Nature*, no seu *site* <http://www.nature.com/nature/debates/e-access/>.

Pesquisadores passaram a criar arquivos eletrônicos de *preprints* e *posprints* como alternativa para publicação direta de seus trabalhos em texto completo, os assim

* Tradução nossa

chamados *open archives*. Este movimento vem crescendo desde então e como resultado deste debate surgem iniciativas como a NCSTRL – Networked Computer Science Technical Report Library (<http://www.ncstrl.org>) –, a NDLTD – Networked Digital Library of Theses and Dissertations (<http://www.ndltd.org>) – e mais recentemente o PUBMED Central (<http://www.pubmedcentral.nih.gov/>) – Public Library of Science (<http://www.publiclibraryofscience.org/>) – e OpenArchives Initiative (<http://www.openarchives.org>).

Os padrões de tecnologia da informação utilizados ou derivados da *Open Archives Initiative* têm um impacto potencial muito grande sobre os sistemas de informação em C&T, afetando substancialmente a maneira como bibliotecas e centros de documentação desempenham funções tradicionais, como seleção, aquisição, registro/tratamento técnico/indexação/classificação e disseminação. Colocam também a questão da cooperação entre sistemas de informação com vistas ao acesso à informação em C&T em um novo patamar. Este trabalho se propõe a discutir este novo quadro que surge com a emergência de documentos digitais em C&T, publicados diretamente na rede e armazenados em bibliotecas digitais e repositórios como os Open Archives, e como estes padrões e tecnologias podem ser usados pelos países em desenvolvimento para ampliar o alcance e o nível de cooperação de seus sistemas de informação.

DOCUMENTOS E ARQUIVOS DIGITAIS EM CIÊNCIA E TECNOLOGIA

A ciência hoje é uma atividade altamente institucionalizada. O conhecimento passa a ter um papel fundamental como insumo produtivo, em um processo que tem suas origens no Renascimento e se acentua fortemente na Revolução Industrial (Gómez, 1987). Sua gestão, a utilização e principalmente o acesso se tornam cada vez mais atividades econômicas críticas (Marcondes, 2001a).

Nesta ciência tão institucionalizada, não existe praticamente lugar para o gênio isolado, capaz de dar conta de uma descoberta científica do início ao fim. A ciência atual é fundamentalmente um trabalho coletivo, em que pesquisadores e grupos de pesquisa trabalham sobre resultados já obtidos por seus pares, e tem como objetivo acrescentar um tijolo a mais em um vasto edifício.

Daí o papel fundamental que desempenha neste contexto a comunicação científica. A ciência não pode avançar sem mecanismos eficientes de comunicação cien-

tífica que integrem em um ciclo a produção de conhecimento, o registro dos resultados, a coleta e estocagem destes registros, a disseminação dos resultados e o reuso, tanto em atividades produtivas quanto como fonte para gerar novos conhecimentos. Até muito recentemente, o periódico científico tradicional era o veículo por excelência para a comunicação científica, mas estamos no meio de uma profunda mudança desta situação (Day, 1999).

Os mecanismos de comunicação científica combinavam canais desde os informais, como conversas pessoais, cartas, telefonemas, *preprints* distribuídos aos pares, bastante ágeis e atualizados em termos de velocidade, mas sem qualquer filtro de qualidade, com canais cada vez mais formais, como as comunicações e trabalhos em congressos e os artigos de periódico, em que os trabalhos para serem aceitos são submetidos à avaliação por pares. No percurso dos canais informais para os formais, a maior rapidez na comunicação é trocada pela maior lentidão da publicação dos resultados de pesquisas e pela maior qualidade graças à revisão por pares (Ziman, 1979; Le Coadic, 1996).

Desde a publicação do *Journal des Savans* e das *Philosophical Transactions da Royal Society*, em 1665 (Day, 1999), que o periódico acadêmico vem tendo um papel destacado como veículo por excelência de comunicação científica. A partir do fim do século XIX, com a instituição do sistema de revisão por pares, este papel se institucionalizou, juntamente com todo aparato acadêmico e científico, atingindo a configuração institucional atual. No mundo dos documentos impressos, os periódicos científicos tinham papel destacado como coroamento de um sistema de comunicação científica institucionalizado e reconhecido pela comunidade acadêmica, que fazia um compromisso entre velocidade e filtros de qualidade. Dentre os papéis por eles cumpridos, destacam-se os seguintes: disseminação ampla dos resultados de pesquisa; controle de qualidade, através do mecanismo de revisão por pares; homologação de prioridade nas descobertas científicas; reconhecimento dos autores; criação de um arquivo público de conhecimentos com cópias armazenadas em bibliotecas de centros de documentação (Day, 1999).

Paralelamente à criação e à institucionalização destes mecanismos de comunicação científica e de filtros de qualidade, vive-se um processo de parcelamento e fragmentação da ciência em áreas cada vez mais especializadas. A isto se soma um processo de identificação e questionamento de novos problemas de pesquisa que de-

mandam o aporte de diferentes áreas do conhecimento como as ciências cognitivas, a ecologia, a moderna visão integrada das ciências da vida etc., levando à identificação de pontos de contato entre diferentes ramos da ciência.

A consequência deste processo para a comunicação científica é a fragmentação e o surgimento de novos periódicos e eventos científicos, dedicados a novas áreas ou a novos problemas de caráter interdisciplinar. Devido à fragmentação descrita, estes veículos de comunicação científica têm audiências cada vez mais restritas e especializadas, levando à diminuição das edições e ao consequente encarecimento do custo final dos periódicos científicos e edição de anais de eventos para seus usuários.

Entretanto, a produção destes veículos de comunicação científica tornou-se um negócio que movimenta vultosas somas e é dominado por grandes empresas, os grandes editores científicos. Os custos das assinaturas de periódicos cada vez mais proibitivos para bibliotecas e centros de documentação, como intermediárias no ciclo de comunicação científica, ameaçam quebrar este ciclo. Este processo vem tendo consequências sérias para o desenvolvimento da ciência e especialmente sérias para países em desenvolvimento, com recursos escassos a serem investidos na ciência e em apoio bibliográfico para a mesma, como o Brasil.

O surgimento da Internet e dos mecanismos de publicação direta na rede tem sido visto pela comunidade acadêmica como uma possível alternativa. O maior retorno que a comunidade acadêmica almeja, publicando os resultados de suas pesquisas, é que estes possam servir de bases a outras pesquisas, sendo citados por outros trabalhos. A citação é a medida clássica do prestígio e do valor de uma contribuição para a ciência em geral. Com base nos mecanismos de citação, são criados indicadores consagrados para cientometria e bibliometria como o *Science Citation Index*, produzido pelo ISI - Institute of Scientific Information (<http://www.isinet.com>). Estudos recentes confirmam que as publicações eletrônicas são muito mais citadas que as publicações em papel: "*The mean number of citations to offline articles is 2,74, and the mean number of citations of online articles is 7.03, an increase of 157%*" (Lawrence, 2001).

À medida que os periódicos científicos se tornam crescentemente mais caros, vem diminuindo sua audiência, resultando disso menor impacto das comunicações nele veiculadas. Segundo este ponto de vista, a cobrança e as restrições ao acesso por parte dos grandes

editores científicos internacionais impediriam o livre fluxo dos resultados da pesquisa e o próprio avanço desta, em prol de interesses comerciais restritos. As consequências deste processo para o desenvolvimento da ciência começam a ser percebidas por um número crescente de cientistas e pesquisadores que se lançaram na busca de alternativas. Stevan Harnard, um dos ideólogos deste movimento, assim formula esta questão:

"Unlike the authors of books and magazine articles, who write for royalty or fees, the authors of refereed journal articles write only for 'research impact'. To be cited and built on in the research of others, their findings have to be accessible to their potential users. From the authors' viewpoint, toll-gating access to their findings is as counterproductive as toll-gating access to commercial advertisements".

E ainda:

"Researchers never benefited from the fact that people had to pay access tolls to read their papers (as subscriptions, and for the online version, site-licences or pay-per-view). On the contrary, those access barriers represent impact barriers for researchers, whose careers and standing depend largely on the visibility and uptake of their research".

Este debate envolvendo pesquisadores e grande editores está sendo coberto pela revista *Nature* no seu site <http://www.nature.com/nature/debates/e-access/>.

A comunidade científica vê as publicações eletrônicas na rede como um meio de aumentar sua visibilidade, acelerar o avanço da ciência e disseminar amplamente os resultados das pesquisas, vistas como patrimônio da humanidade (Harnard). Ela soube, com muita propriedade, acercar-se das novas possibilidades abertas pelas tecnologias de informação para criar mecanismos alternativos de publicação de resultados das atividades de pesquisa.

Observa-se, nos últimos anos, uma forte tendência, surgida no seio da comunidade científica mundial, para a criação de arquivos eletrônicos informais e autogeridos, voltados para este fim. Estes arquivos são conhecidos como *eprints archives* e configuram claramente uma transição do modelo de comunicação tradicional, baseado em publicações periódicas formalmente estabelecidas, para um novo e surpreendente paradigma. Eles exemplificam o mais democrático e eficiente modelo para disseminação de resultados de pesquisa. Uma referência importante em português sobre o histórico, os desenvolvimentos recentes e a dimensão da OpenArchives

Initiative, inclusive com uma entrevista a Paul Ginsparg, pode ser encontrada em Sena (2000).

Além das condicionantes tecnológicas oferecidas pela Internet, que facilitaram extraordinariamente a publicação de documentos eletrônicos, há um consenso absoluto de que outros fatores foram decisivos na emergência dos arquivos de *eprints*, principalmente os seguintes: a) a lentidão do ciclo de edição das revistas comparado à rapidez da geração de novos conhecimentos de algumas áreas; b) a renúncia ao direito sobre a obra imposto pelas revistas, que impede a ampla disseminação pelo autor de suas idéias; c) a perspectiva extremamente rígida e conservadora dos esquemas de *peer review* adotados pelas revistas, que não raro são impeditivos ao surgimento de idéias inovadoras, como é ilustrado no artigo de Martin (1993) sobre uma hipótese pouco ortodoxa sobre o surgimento da Aids, cuja publicação foi recusada pelos principais periódicos do *establishment* médico; d) o alto custo da subscrição dos periódicos, seja em papel ou em meio eletrônico. Os arquivos abertos não são, como possa parecer a princípio, uma proposta anárquica, que elimina os critérios de qualidade da ciência; muitos deles utilizam o esquema de *peer-review*, ou uma separação entre textos avaliados e não-avaliados; muitos destes textos são cópias livres de artigos já publicados ou a serem publicados em periódicos convencionais (Harnard).

Muito recentemente, em 1999, em um passo seguinte à criação dos arquivos *eprints*, a comunidade científica internacional se mobilizou para torná-los *interoperáveis*, isto é, passíveis de serem consultados *simultaneamente*. Esta interoperabilidade foi alcançada mediante adoção de um conjunto de especificações técnicas e princípios organizacionais bastante simples, porém potencialmente poderosos e de grande alcance, no objetivo de integração desses arquivos. Esta iniciativa é conhecida como OAI – Open Archive Initiative, <http://www.openarchives.org/> (Van de Sompel, 2000) e tem como objetivos básicos apoiar o desenvolvimento de arquivos de *eprints* e criar uma arquitetura tecnológica padronizada que sustente a *interoperabilidade* entre eles. No bojo da OpenArchives Initiative, foram desenvolvidas tecnologias, padrões e metodologias para publicação, disponibilização, metadados e intercâmbio automático de metadados entre bibliotecas digitais. A dimensão da iniciativa dos Open Archives em nível mundial e sua cobertura regional e temática podem ser mais bem avaliadas consultando-se a lista dos arquivos eletrônicos existentes em <http://www.osti.gov/eprints/ppnbrowse.html>.

Um repositório eletrônico aberto apresenta características específicas. Seu *site* apresenta facilidades que permitem a um autor submeter diretamente seus trabalhos, armazená-los em forma digital permanentemente, editá-los, substituí-los e receber críticas e contribuições; ao submeter um trabalho para armazená-lo e disponibilizá-lo no arquivo eletrônico, um autor também o descreve, em um formulário de catalogação, de onde serão extraídos os metadados como autor, título, idioma, assunto etc. que permitirão recuperar o documento; os metadados são, portanto, obtidos como um subproduto da submissão de um documento. O *site* permite também a consulta e o acesso direto aos trabalhos eletrônicos nele armazenado. Um servidor de *eprints* compatível com o OpenArchives Initiative Protocol for Metadata Harvesting – OAI PMH – permitirá a exposição de metadados dos trabalhos nele armazenados para coleta automática (*harvesting*) e reuso por provedores de serviços de informação, que com eles podem criar diferentes serviços de valor agregado.

A publicação na rede de textos completos de interesse para C&T começa a ser uma realidade também no Brasil. Iniciativas pioneiras como o SCIELO – <http://www.scielo.br> – (Packer, 1998), um portal que abrange dezenas de periódicos, associado a uma metodologia para publicar e prover acesso a periódicos eletrônicos em texto completo, os diferentes periódicos brasileiros já publicados na *web* em texto completo, a proposta de desenvolvimento de um ambiente *web* para edição de anais de congresso pelo CIN/CNEN, as recentes iniciativas do arquivo de *eprints* do IMPA (Chataignier, 2001), as publicações digitais da biblioteca digital do LAMBDA – Laboratório de Automação de Museus, Bibliotecas Digitais e Arquivos – da PUC-Rio – <http://www.lambda.maxwell.ele.puc-rio.br>, o Banco de Teses e Dissertações do Programa de Pós-graduação em Engenharia de Produção da UFSC – <http://teses.eps.ufsc.br/> e do repositório de teses da USP – <http://www.teses.usp.br> –, (Masiero, 2001) mostram que esta é uma tendência irreversível aqui também.

Recentemente (Marcondes, 2001b) o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), por meio do projeto da Biblioteca Digital Brasileira em C&T (BDB), passou a fomentar o desenvolvimento de recursos informacionais brasileiros de interesse para C&T em texto completo, como teses e dissertações, artigos de periódicos, trabalhos em congressos, arquivos eletrônicos de *preprints*, integrando e provendo interoperabilidade entre estes recursos mediante acesso unificado aos mesmos, via única interface *Web*. O projeto prevê o uso do

OAI PMH como um dos mecanismos para prover interoperabilidade entre os diversos recursos informacionais contendo documentos digitais de interesse para C&T, coletando seus metadados para uma base comum de onde será feito o acesso.

COOPERAÇÃO ENTRE SISTEMAS DE INFORMAÇÃO NA ERA DOS DOCUMENTOS E ARQUIVOS DIGITAIS

A produção da ciência e a comunicação dos resultados de pesquisa na nossa sociedade se dão de forma dispersa, segundo dois eixos: *dispersa espacialmente* – pesquisadores em diversos locais geográficos produzem e comunicam seus resultados de pesquisa por mecanismos de comunicação científica diferentes; *dispersa temporalmente* – pesquisadores produzem e comunicam seus resultados em momentos diversos.

Neste contexto, os sistemas de informação desempenham um papel fundamental na economia da informação-conhecimento, agregando valor ao servirem como pontos focais, que concentram a informação científica, produzida, por natureza, de forma dispersa. A eles os usuários recorrem para encontrar as informações de que necessitam, com um alto grau de probabilidade de encontrá-las.

Tradicionalmente a informação de interesse para a pesquisa científica é em grande parte composta pela chamada documentação não-convencional, também chamada de “literatura cinzenta”, documentos que não são encontrados no circuito editorial convencional, como relatórios de pesquisa, trabalhos apresentados em eventos, *preprints*, teses e dissertações, que noticiam com grande atualidade os resultados de pesquisa.

O papel dos sistemas de informação sempre foi o de se contrapor a esta dispersão, provendo um ponto de concentração para a comunicação dos resultados das pesquisas. Os sistemas especializados em ICT organizaram-se para prover acesso principalmente a este tipo de documentação, a “literatura cinzenta”, como o NTIS, o Dissertation Abstracts, o Chemical Abstracts, INIS, AGRIS, ERIC e muitos outros. Estes sistemas trabalham dentro do paradigma organizacional/metodológico/tecnológico da informação referencial como etapa para o acesso a cópias de documentos em papel.

Coletar esta “literatura cinzenta” sempre foi caro e extremamente trabalhoso para os sistemas de ICT, devido à dispersão da literatura científica. *O surgimento das publicações eletrônicas começa a mudar radicalmente este qua-*

dro. Hoje em dia não é mais suficiente para garantir o máximo de visibilidade de seu acervo que bibliotecas digitais simplesmente disponibilizem seus dados na Internet. A quantidade de informações disponível na rede é tão grande que identificar, localizar, descobrir a existência e acessar informações relevantes torna-se um problema crítico, demandando um tempo proibitivo aos usuários.

Às bibliotecas digitais é novamente colocada, como já foi há tempos para as bibliotecas convencionais, a questão de cooperarem, agora sob novas bases organizacionais e tecnológicas, para garantir o máximo de visibilidade a seus acervos. Atingir esta visibilidade não significa mais necessariamente que alguém buscando informações terá de acessar o *site* da biblioteca digital ou arquivo eletrônico para ter acesso aos documentos digitais nele depositados. A possibilidade que seus acervos possam ser consultados simultaneamente, sem que um usuário acesse cada *site* individualmente, a chamada *interoperabilidade*, tem sido perseguida como um mecanismo que viabilize esta possibilidade. Atingir a interoperabilidade entre repositórios de *eprints* ou bibliotecas digitais, distintos e heterogêneos, possibilitando que possam ser consultados simultaneamente, envolve um aporte intenso em termos de tecnologias, protocolos e padronização.

O OAI PMH é um protocolo que provê interoperabilidade não imediata (ou seja, não é, portanto, um protocolo para busca *on-line*) entre repositórios de *eprints*, bibliotecas digitais ou qualquer servidor na rede que queira *expor*, ou seja, tornar visíveis metadados de documentos nele armazenados para um programa externo que queira *coletá-los*. Os participantes da OpenArchives Initiative rejeitaram opções como, por exemplo, o protocolo de recuperação de informações Z39.50, que distribui uma busca imediata e simultânea por vários servidores que hospedam catálogos de bibliotecas capazes de resolvê-la, em favor de uma solução mais simples e menos onerosa em termos de recursos computacionais consumidos (Troll, 2001). As diferenças na interoperabilidade com o uso do protocolo Z39.50 X o OAI-PMH podem ser bem visualizadas na figura 1, a seguir.

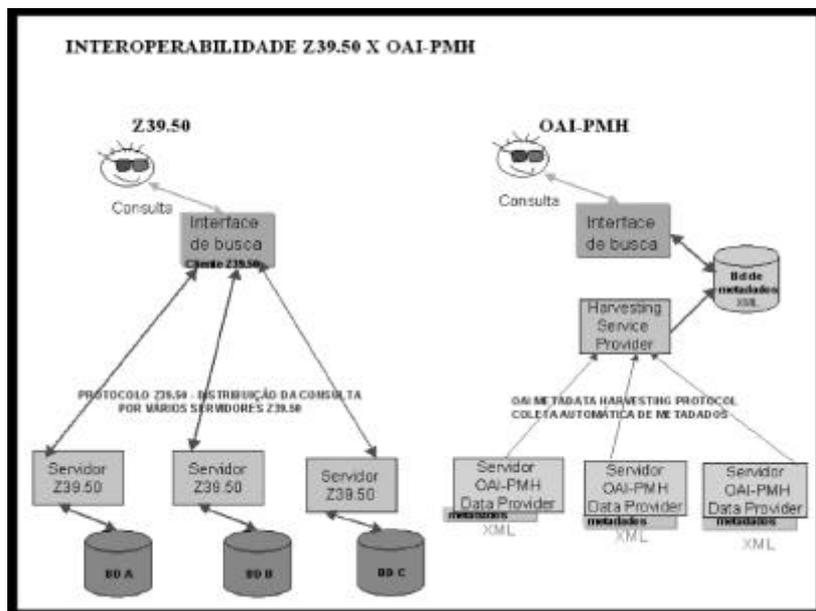
O OAI PMH define a troca de solicitações e de metadados entre o servidor de *eprints* e um programa robô externo. Dentro da concepção OAI, existem as instituições chamadas provedoras de dados (*Data Providers*), que são bancos de documentos eletrônicos que oferecem facilidades para publicação e armazenamento de documentos eletrônicos e sua disponibilização em um servidor conectado à Internet, e as instituições provedoras de serviços

(Service Providers), que coletam metadados de um ou mais provedores de serviço e com estes metadados prestam serviços de valor agregado. Exemplos destes serviços seriam o acesso unificado a acervos de diferentes provedores de dados, por meio de um portal *web* único ou a constituição de bases de dados qualificadas sobre temas específicos, ou um periódico eletrônico com textos avaliados e submetidos a um esquema de *peer-review*, desenvolvido a partir dos metadados coletados de diversos provedores de dados.

A troca de mensagens entre o servidor do provedor de dados e o programa robô externo do provedor de serviços para a transferência de metadados é unidirecional – o provedor de serviços faz solicitações ao provedor de dados, que responde enviando metadados. As solicitações do provedor de serviço são feitas via protocolo HTTP (<http://www.ietf.org/rfc/rfc2616.txt>), usando comandos CGI* codificados por meio dos métodos GET ou POST. As solicitações são respondidas pelo provedor de dados com o envio de dados das respostas ou metadados dos documentos armazenados, codificados em XML (<http://www.w3.org/XML/>). O OAI-PMH estabelece o Dublin Core Metadata Element Set (<http://purl.org/DC/documents/rec-dces-19990702.htm>) como conjunto mínimo de metadados a ser suportado pelos provedores de dados em resposta a uma solicitação de um provedor de serviços. No entanto, o provedor de serviços pode, a seu critério, oferecer outros formatos de metadados, mais amplos e complexos, como o MARC.

Os metadados, formando um registro de cada documento armazenado no provedor de dados, têm um identificador único, formado pelo identificador do provedor de dados mais um identificador do registro. Cada registro apresenta também um selo temporal, denominado *datestamp*, que indica a data da criação ou última alteração do documento associado a este registro; o *datestamp* é a chave que permite a coleta automática dos metadados do provedor de dados a partir de uma determinada data, possibilitando, portanto, a sincronização entre os registros do provedor de dados e de um provedor de serviços que forneça um serviço de acesso simultâneo

FIGURA 1
Interoperabilidade Z39.50 X OAI-PMH



a metadados de documentos armazenados em diversos provedores de serviço.

O protocolo prevê ao todo seis “verbos” que um programa robô de provedor de serviços pode enviar a um provedor de dados para coletar metadados de documentos aí armazenados: *identify* – obtém dados administrativos sobre o provedor de dados, sua política de publicação de documentos, seu escopo etc.; *ListSets* lista as classificações sob as quais os documentos são organizados no provedor de dados; *ListMetadataFormats* – lista os formatos de metadados por meio dos quais os metadados dos documentos armazenados no provedor de dados podem ser apresentados; *ListIdentifiers* lista os identificadores de registros armazenados no provedor de dados, podendo opcionalmente limitar estes registros a partir de uma data, ou pertencentes a um *set*; *ListRecords* lista os metadados dos registros armazenados no provedor de dados segundo um formato de metadados, especificando todos que pertencem a um *set* ou todos a partir de uma data; e *GetRecords* obtém os metadados dos registros armazenados segundo um formato de metadados, dado um identificador de registro. A interação entre Service Providers e Data Providers usando o OAI-PMH pode ser visualizada na figura 2, a seguir.

A interação entre provedores de dados e provedores de serviços usando o OAI-PMH pode ser vista através do OAI Repository Explorer, uma interface interativa para testar a compatibilidade de repositórios de documentos

* CGI é Common Gateway Interface, um padrão entre formulários HTML em um computador cliente e programas em um servidor.

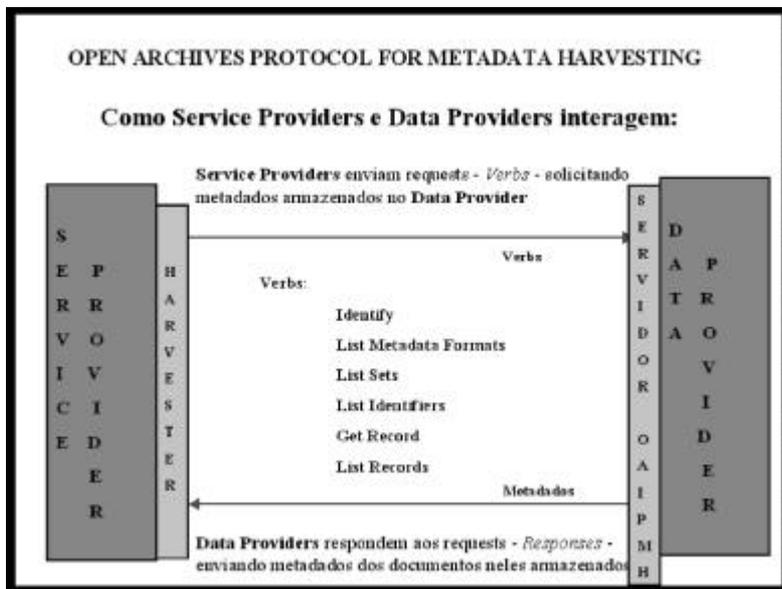
eletrônicos com OAI PMH, no site <http://rocky.dlib.vt.edu/oai/cgi-bin/Explorer/oai1.0/testoai>.

NOVOS SERVIÇOS COOPERATIVOS VIABILIZADOS PELO OAI PMH

Os países em desenvolvimento têm dificuldades históricas, não só no acesso aos resultados de pesquisas, mas também em dar visibilidade à ciência por eles produzida, não obstante ser uma ciência de grande importância para uma parcela imensa da população mundial (exemplo mais eloqüente disto são as pesquisas da Fiocruz que levaram à produção de várias substâncias que fazem parte do coquetel anti-*aids* que podem ser um alívio para a tragédia africana). Isto acontece principalmente por conta da série de impeditivos encontrados para a inserção do produto dessa ciência nos sistemas de informação e nas bases de dados internacionais. Esses impeditivos são de natureza distinta, tais como idioma (tudo que não é publicado em inglês é considerado hieróglifo), publicação em periódicos não coletada pelas bases de dados internacionais, o laboratório que sedia a pesquisa está fora do circuito da *big science*, o não-enquadramento ao *main stream* da ciência mundial e assim por diante (Sayão, 1996). Assim sendo, é de grande importância a criação de sistemas regionais, como o LILACS, que além de prestar serviços de informação aos seus usuários em todo o mundo, estabelecem meios de disseminação adequados à sua audiência e contribuem para incluir a ciência do Terceiro Mundo nos fluxos internacionais.

Informação em ciência e tecnologia livre na Internet, associada ao conjunto de metodologias colocadas à disposição da comunidade acadêmica pela Open Archive Initiative (OAI), abrem grandes possibilidades para os sistemas de informação que se dispuserem a avaliar com espírito criativo as oportunidades oferecidas por estas metodologias. Uma série de novos serviços baseados em reuso de metadados pode ser concebida, incluindo redes cooperativas e sistemas de informação regionais. O conjunto de metodologias OAI PMH – protocolo de coleta automática de metadados – é um protocolo de fácil implementação. Outro dado promissor é o que indica ser possível montar repositórios digitais Open Archives em servidores PC-Intel, rodando sistema operacional Unix/Linux com *softwares* que podem ser obtidos livremente

FIGURA 2
Interação Service Providers e Data Providers



da Web (<http://www.eprints.org>). Estes repositórios serão compatíveis com o número crescente de repositórios digitais espalhados por todo mundo, nas mais diversas áreas, permitindo alto grau de interoperabilidade entre eles.

Nesta direção, as instituições desses países podem se aproveitar das metodologias OAI para expor os seus metadados aos repositórios digitais internacionais e assim dar mais visibilidade aos seus conteúdos; podem também reaproveitar os metadados disponíveis em escala planetária pelos repositórios digitais OAI para coleta automática, agregando valores e criando versões novas para os serviços e produtos de informação já tradicionais, como busca retrospectiva e DSI e, sobretudo, criando novos conceitos de serviços próprios para o ambiente de rede, como é o caso das bibliotecas servindo como publicadoras Web dos trabalhos de sua comunidade de usuários e dos esquemas de acompanhamento das novidades publicadas na Internet em áreas específicas e da descoberta de recursos informacionais.

A proposta de maior alcance potencial da OAI é a que diz respeito à integração de repositórios digitais livres por meio de serviços independentes. Esta proposta indica um caminho viável para a construção de uma estrutura global unificada para a literatura acadêmica, que inclui não somente *preprints*, mas também periódicos, relatórios, anais de congresso e outros tipos de literatura acadêmica. Em um outro eixo, esta estrutura global cria também facilidades para o surgimento de sistemas de in-

formação organicamente mais complexos, como são as redes cooperativas locais, regionais ou internacionais. O conjunto de metodologias e protocolos OAI torna o ônus administrativo-financeiro destas redes muito menor, a tarefa de criar os catálogos coletivos mais simples e, sobretudo, deixa os centros cooperantes com um grau maior de independência, principalmente para cooperar com outros serviços/sistemas aderentes ou não ao protocolo OAI. Por exemplo: poderíamos ter uma rede de teses eletrônicas das universidades paulistas que estivesse integrada ao Consórcio Brasileiro de Teses Eletrônicas, que, por sua vez, fosse cooperante, ao mesmo tempo, de uma rede de teses em língua portuguesa, formada pelos países que falam este idioma e também de uma rede de teses eletrônicas de países latino-americanos.

No entanto, para se chegar a esse nível de integração e articulação entre redes e sistemas de ICT, é necessária a criação de níveis distintos de interoperabilidade (Miller, 2000), que se superponham à interoperabilidade tecnológica estabelecida pelo conjunto de padrões e protocolos preconizados pelo OAI e viabilizados facilmente pelo uso do *software Eprint*.

A interoperabilidade tem muitas faces: é ela que permite que sistemas de ICT distintos e heterogêneos possam aproveitar e agregar valor à informação criada por outro, gerando novos serviços e novas visões para a mesma informação. Não é, entretanto, uma questão meramente técnica e tecnológica, ela também depende da gestão, articulação e cooperação mútua entre sistemas de ICT no plano político. Esta faceta da interoperabilidade, aqui chamada de “interoperabilidade política”, depende fundamentalmente da criação de organizações detentoras de canais e fóruns adequados, nos quais a discussão e o consenso possam se estabelecer e as decisões possam ser tomadas endossadas pelo grau de representatividade dessas organizações. Este é o caso da DLF (Digital Library Federation), da PILA (Publisher International Linking Association) e do próprio OAI (Open Archives Initiative), para citar somente três iniciativas internacionais importantes neste momento.

Outra faceta importante do mesmo problema é a “interoperabilidade semântica”. Ela requer o uso generalizado de instrumentos comuns de descrição temática. Um conjunto mínimo de metadados padronizados, uma linguagem de descrição temática de cobertura ampla, tal como a Tabela de Áreas do Conhecimento Capes/CNPq e a criação de *servidores de autoridades* cooperativos. A interoperabilidade semântica tem como reflexo imediato a melhoria na qualidade da recuperação das infor-

mações e a otimização da consulta a sistemas interligados. Esses resultados são bastante perceptíveis pelo usuário final.

A Internet nos traz outra forma de interoperabilidade que é a *linkagem* ou enlaces entre sistemas. A *linkagem* permite a navegação via *hiperlinks* entre as várias manifestações do trabalho acadêmico de um indivíduo, normalmente dispersas em vários sistemas, seja como autor, orientador ou membro de banca de teses ou dissertações eletrônicas, seja como autor de artigos de periódico, de trabalhos em congressos ou acessando seu currículo em um sistema de currículos. Mas é de fundamental importância a adoção de padrões e metodologias que garantam a persistência dos endereços eletrônicos dos recursos informacionais, como o PURL - Persistent URL (<http://www.purl.org/>) e o DOI - Digital Object Identifier (<http://www.doi.org/>), no sentido de preservar o investimento na *linkagem* entre sistemas.

A seguir são descritos alguns produtos e serviços de informação que podem ser implementados usando-se as metodologias OAI.

– **Formação de bases de dados**

A alternativa de gerar bases de dados a partir de informações disponíveis na Internet e do reuso de seus metadados se torna cada vez mais uma opção viável que começa a ser adotada por sistemas internacionais de grande porte como é o caso do Energy Technology Data Exchange (ETDE), da Agência Internacional de Energia. Isto acontece principalmente pelo alto custo da coleta manual e do tratamento técnico - catalogação e indexação -, fatores que têm inviabilizado a sustentação econômica do modelo tradicional de formação de bases de dados que, diga-se de passagem, sempre foi deficitário e precisava de fortes subsídios externos. A coleta automática de metadados - *harvesting* -, baseada no protocolo OAI, pode facilitar tremendamente a formação de bases de dados, diminuir drasticamente os custos de coleta, catalogação e indexação e também o custo de gestão destes processos. E o mais importante: as bases de dados organizadas dessa forma têm agregado um valor importantíssimo que é o acesso ao objeto digital referenciado, seja ele um texto, um vídeo ou uma peça de museu. Pode-se construir bases temáticas, qualificadas, por tipologia de documento (teses, relatórios), orientadas por problemas (epidemia de dengue, crise energética), orientadas por projetos ou missão (prospecção de petróleo em águas profundas, despoluição da Baía de Guanabara), com metadados e informações disponíveis por todo mundo.

– **Estabelecimento de redes cooperativas nacionais, regionais e internacionais**

O protocolo OAI pode facilitar sobremaneira o estabelecimento de redes de informação. Um exemplo importante é a rede de teses eletrônicas da Virginia Tech. University – EUA –, a NDLTD que tem alcance mundial. No Brasil, dentro do âmbito da BDB, tem-se o Consórcio Brasileiro de Teses Eletrônicas, projeto em andamento que, inicialmente, reúne as teses e dissertações eletrônicas da USP, UFSC, PUC-Rio, ENSP (Fiocruz) e conta com o apoio do CNPq. O Consórcio tem a perspectiva de estender-se por Portugal, por meio de convênio com a Universidade do Minho. O uso do Protocolo OAI pode ainda tornar-se uma alternativa a mais para as redes já estabelecidas, criando novas opções de contribuição e coleta menos onerosa em termos financeiros e gerenciais para os membros dessas redes. Existe uma série de *software* livres, do tipo *front-end*, que podem tornar uma base de dados de metadados aderente ao protocolo OAI sem interferências com os processos internos da rede. Muitos destes *software* podem ser encontrados em <http://www.openarchives.org/tools/tools.htm>. É também possível estabelecer formas ainda mais simples de coleta automática baseadas, por exemplo, na transferência de arquivos-textos ou arquivos HTML, adequadamente formatados, via FTP ou via mecanismos simples de coleta automática, que podem implementar formas simples de crítica de dados.

– **Edição de periódicos eletrônicos**

O *software* Eprint (distribuído livremente) desenvolvido pela Universidade de Southampton – UK, para disseminar a implantação de repositórios digitais compatíveis com OAI, oferece facilidades interessantes para quem deseja editar revistas diretamente na rede. A característica mais marcante para esta aplicação é que o *software* implementa um ambiente de submissão eletrônica bastante sofisticado, que permite avaliação por parte de revisores e comentários vindos dos leitores, além de instrumentos para a gerência de publicação dos artigos em uma biblioteca digital. Está crescendo rapidamente o número de revistas eletrônicas publicadas com esta tecnologia, e isto indica mais um serviço que será descrito a seguir.

– **Catálogo coletivo de revistas eletrônicas OAI**

Já é possível implementar catálogos coletivos de revistas eletrônicas que tenham seus metadados coletados com o uso do Protocolo OAI, que permitam busca simultânea em todas as revistas e a recuperação dos artigos em texto

completo. Esses catálogos podem ser enriquecidos com *links*, classificados por áreas do CNPq, por exemplo, para o universo em expansão das revistas eletrônicas livres na Internet, tornando-se uma opção extremamente barata às assinaturas dos periódicos convencionais, sejam eles impressos ou eletrônicos.

– **Ambiente para submissão eletrônica de contribuições a congressos**

O Centro de Informações Nucleares está no momento desenvolvendo, também dentro do escopo da BDB, o projeto Edição Eletrônica de Anais de Eventos em C&T, que também fará uso das facilidades de submissão eletrônica e de críticas oferecidas pelo *software* Eprint. O projeto prevê a criação de uma biblioteca digital de anais de congressos, além de mecanismos para edição de anais impressos por meio convencional ou/e em meio digital.

– **Biblioteca como publicadora Web**

Nesta nova configuração, a biblioteca e os demais unidades de informação podem assumir um papel de grande relevância que é o de publicador *Web* da produção acadêmica de sua comunidade de usuários. O investimento para isso é pequeno: é necessário equipá-la com servidores compatíveis com o Protocolo OAI, que podem ser microcomputadores PC-Intel ligados à Internet, rodando o conjunto de *software* que formam o Eprints. É ainda necessário o estabelecimento de políticas de submissão e metodologias de gestão dos arquivos digitais. Um dado importante é que unidades de informação que possuem estes servidores serão interoperáveis entre si, podendo trocar dados via processos de *harvesting* automático ou via processos tradicionais. Isto traz a perspectiva de outros serviços descritos a seguir.

– **Busca unificada**

Um bom exemplo é o Arc – Cross Archive Searching Service – primeiro serviço a proporcionar acesso integrado a diversos arquivos eletrônicos – 30 em março de 2001 – a partir de uma única interface <http://arc.cs.odu.edu/>. Outro exemplo é o *site* My.OAI, <http://www.myoai.com/>.

– **Descoberta de recursos via perfis de DSI**

É possível implementar via processos de *harvesting*, a descoberta de recursos informacionais em fontes específicas, previamente selecionadas. As informações capturadas podem ser filtradas segundo perfis de DSI, funcionalidade já implementada pelo *software* Eprints, e enviadas via *e-mail* para indivíduos, comunidades, departamen-

tos, programas etc., dependendo do nível de capilaridade que se queira alcançar.

Outros exemplos de serviços baseados em “harvesting” de metadados usando o OAI PMH, implementados pela DLF – Digital Library Federation, <http://www.diglib.org/> - muitos já em caracter operacional, estão relatados em Waters (2001).

CONCLUSÕES

Só aos poucos a comunidade envolvida com publicações eletrônicas e bibliotecas digitais vem conhecendo e avaliando o grande potencial de interoperabilidade do OAI PMH. Nestas comunidades, o protocolo vem tendo grande aceitação, devido provavelmente à sua grande simplicidade conceitual e tecnológica. Já existem propostas de extensões para este protocolo, endereçando as interfaces entre módulos que comporiam uma biblioteca digital completa (Suleman, 2001). Maiores detalhes técnicos sobre o OAI PMH podem ser encontrados em Lynch (2001) e Warner (2001).

A divisão de trabalho entre provedores de dados e provedores de serviços prevista pelo OAI PMH, possibilitando a criação de novos serviços de valor agregado, abre grandes possibilidades. Nada impede que uma biblioteca digital seja simultaneamente provedor de dados e de serviços. Na era dos documentos em papel, cada cópia deste documento, armazenada em uma biblioteca, era descrita e indexada, produzindo registros que, organizados em fichários manuais ou catálogos automatizados, eram mecanismos para viabilizar a recuperação destes documentos. Portanto, a criação de metadados era separada da produção do documento e repetida em diferentes bibliotecas. A criação de metadados simultaneamente à submissão de um documento digital para publicação em um arquivo eletrônico, isto é, o encapsulamento do documento e seus metadados, realiza na prática a proposta antiga na área de biblioteconomia, da catalogação na fonte. Esta se torna viável pelo uso de um conjunto de metadados – o Dublin Core – tão simples “a ponto de poder ser utilizado pelo próprio autor” (Weibel, 1995). A biblioteca fica dispensada de realizar o tratamento técnico deste documento, geralmente tão oneroso. O documento tratado na fonte se torna imediatamente insumo para diferentes serviços, possibilitando seu reuso. A criação de documentos digitais formatados em XML (Extensible Markup Language (XML), 2000) avança ainda mais nesta direção: conteúdo e metadados vão compor um todo único, já que a linguagem XML permite a marcação das partes que compõem um documento digital, permitindo que um programa possa identificar que

porção do conteúdo do documento constitui o título, identifica o autor, o assunto, as hipóteses, a conclusão etc. (Berners-Lee, 2001).

A possibilidade de coleta automática de metadados viabilizada pelo OAI PMH é a chave para uma nova prática de cooperação entre bibliotecas, desonerando os cooperantes do pesado ônus administrativo de gerenciar o envio de lotes, correções, atualizações, operações estas que permitem, em um esquema de cooperação tão conhecido dos sistemas de informação brasileiros, manter um catálogo coletivo. *Software* para tornar provedores de dados e de serviços compatíveis com o OAI PMH está disponível gratuitamente na Internet, no *site* da Open Archives Initiative.

No contexto das bibliotecas especializadas, atividades como seleção e aquisição podem tornar-se quase automáticas, realizando periodicamente a coleta automática de metadados dos arquivos eletrônicos suscetíveis de atender à determinada comunidade. Metadados dos documentos digitais coletados já vêm tratados. Uma biblioteca digital não precisa mais ter a posse (Lancaster, 1994) dos documentos que compõem o seu acervo, mas simplesmente manter metadados dos mesmos apontando para o texto completo, armazenado em um outro servidor. O reuso de metadados que também foi sempre perseguido pelas bibliotecas, via sistemas e bancos de catalogação cooperativa, torna-se uma possibilidade real.

O antigo ciclo de comunicação científica, incluindo pesquisador, editor, serviços de informação e biblioteca nas etapas de produção de conhecimentos, registro deste conhecimento, publicação, seleção, aquisição, descrição, armazenamento e disseminação, fica profundamente alterado pelo aporte das tecnologias de informação. Estas etapas se integram, estreitam-se. Produzir textos digitais já implica, praticamente, publicá-los, descrevê-los e disponibilizá-los para disseminação imediatamente.

A existência de documentos livres com os resultados de pesquisas de ponta em diversas áreas de C&T disponíveis na Internet configura uma oportunidade altamente significativa e até então inédita para a ciência dos países em desenvolvimento como o Brasil, conforme analisa Chan (2001). Constitui um mecanismo potencial de democratização no acesso aos resultados de pesquisas e do conhecimento em geral. Esta oportunidade e suas potencialidades não podem passar despercebidas pela comunidade acadêmica brasileira, nem pelos gestores e planejadores de C&T, nem pelos gestores dos sistemas de ICT.

Artigo aceito para publicação em 19-11-2002

REFERÊNCIAS

- BERGMAN, Michael K. The deep web: surface hidden value. *Journal of Electronic Publishing*, v. 7, n. 1, Aug. 2001. Disponível em: < <http://www.press.umich.edu/jep/07-01/bergman.html>> . Acesso em: 15 jun. 2002.
- BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web. *Scientific American*, New York, n. 5, May 2001. Disponível em: < <http://www.scian.com/2001/0501issue/0501berners-lee.html>> . Acesso em: 24 maio 2001.
- CHAN, Leslie, KIRSOP, Barbara. Open archiving opportunities for developing countries: towards equitable distribution of global knowledge. *Ariadne*, v.30, Dec. 2001. Disponível em: < <http://www.ariadne.ac.uk/upcoming—issue31/oai-chan/intro.html>> . Acesso em: 20 mar. 2002.
- CHATAIGNIER, Maria Cecília Pragana; SILVA, Margareth Prevot. Biblioteca digital: a experiência do IMPA. *Ciência da Informação*, Brasília, v. 30, n. 3, p. 7-12, set./dez. 2001.
- DAY, Michael. The scholarly journal in transition and the PubMed Central proposal. *Ariadne*, v. 21, Sept.1999. Disponível em: < <http://www.ariadne.ac.uk/issue21/pubmed/>> . Acesso em: 10 mar. 2002.
- EXTENSIBLE markup language (XML). World Wide Web Consortium, 2000. Disponível em: < <http://www.w3.org/XML/>> . Acesso em: 30 jan. 2001.
- GINSPARG, P. Winners and losers in the global research village. In: CONFERENCE ON ELECTRONIC PUBLISHING IN SCIENCE, 1996, Paris. *Proceedings...* Disponível em: < <http://xxx.lanl.gov/blurb/pg96unesco.html>> . Acesso em: 05 out. 2001.
- GOMEZ, Maria Néida González de. O papel do conhecimento e da informação nas formações políticas ocidentais. *Ciência da Informação*, Brasília, v.16, n. 2, p. 157-167. jul./dez. 1987.
- HARNARD, Stevan. *The self-archiving initiative: nature web debates*. Disponível em: < <http://www.nature.com/nature/debates/e-access/Articles/harnard.html>> . Acesso em: 3 nov. 2001.
- LANCASTER, F. W. Ameaça ou oportunidade? O futuro dos serviços bibliotecários à luz das inovações tecnológicas. *Revista de Biblioteconomia da UFMG*, v. 23, n. 1, p. 7-27, jan./jul. 1994.
- LAWRENCE, Steve. *Free online availability substantially increases a paper's impact: nature web debates*. Disponível em: < <http://www.nature.com/nature/debates/e-access/Articles/lawrence.html>> . Acesso em: 10 jun. 2001.
- LE COADIC, Yves-Francois. *A ciência da informação*. Brasília : Briquet de Lemos, 1996.
- LYNCH, Clyfford. Metadata harvesting and the open archives initiative. *ARL Bimonthly Report*, n. 217, p.1-9, Aug. 2001. Disponível em: < <http://www.arl.org/newsltr/217/PMH.html>> . Acesso em: 18 nov. 2001.
- MASIERO, Paulo Cesar *et al.* A biblioteca digital de teses e dissertações da Universidade de São Paulo. *Ciência da Informação*, Brasília, v. 30, n. 3, p. 34-41, set./dez. 2001.
- MARCONDES, Carlos Henrique. Tecnologias de informação e impacto na formação do profissional de informação. *Transinformação*, Campinas, v. 11, n. 3, p. 189-194, 1999.
- _____. Representação e economia da informação. *Ciência da Informação*, Brasília, v. 30, n. 1, p. 61-70, jan./abr. 2001.
- _____. SAYÃO, Luis Fernando. Integração e interoperabilidade no acesso a recursos informacionais em C&T: a proposta da Biblioteca Digital Brasileira. *Ciência da Informação*, Brasília, v. 30, n. 3, p. 24-33, set./dez. 2001. Disponível em: < <http://www.ibict.br/cionline/300301/3030401.htm>> . Acesso em: 20 dez. 2001.
- MILLER, Paul. Interoperability. What is it and why should I want it? *Ariadne*, v. 24, Jun. 2000. Disponível em: < <http://www.ariadne.ac.uk/issue24/interoperability/intro.html>> . Acesso em: 20 mar. 2002.
- NEC RESEARCH INSTITUTE. Finding a needle in the web. *Boletim Edupage*, Apr. 1998.
- NEWS: Open Archives Initiative Protocol for Metadata Harvesting: general. Version 2. Disponível em: < OAI-general@oaisrv.nsd.cornell.edu> . Acesso em: 8 fev. 2002.
- OPEN ARCHIVES INITIATIVE PROTOCOL FOR METADATA HARVESTING. Version 1.0. Disponível em: < <http://www.openarchives.org/OAI/openarchivesprotocol.html>> . Acesso em: 29 jan. 2001.
- PACKER, Abel *et al.* SciELO: uma metodologia para publicação eletrônica. *Ciência da Informação*, Brasília, v. 27, n. 2, 1998.
- ROBREDO, Jaime, CUNHA, Murilo Bastos. *Documentação de hoje e de amanhã: uma abordagem automatizada da biblioteconomia e dos sistemas de informação*. 2. ed. São Paulo : Global, 1994.
- SAYÃO, Luis Fernando. Bases de dados: a metáfora da memória científica. *Ciência da Informação*, Brasília, v. 25, n. 3, 1996.
- SENA, Natália Kneipp. Open archives: caminho alternativo para a comunicação científica. *Ciência da Informação*, Brasília, v. 29, n. 3, p. 71-78, set./dez. 2000.
- SHNEIDERMAN, Ben. Clarifying search: a user-interface framework for text searches. *Dlib Magazine*, Jan. 1997. Disponível em: < <http://www.dlib.org/dlib/january97/retrieval/01sneiderman.html>> .
- SOMPTEL, Herbert van de; LAGOZE, Carl. The Santa Fe Convention of the Open Archives Initiative. *Dlib Magazine*, v. 6, n. 2, Feb. 2000. Disponível em: < <http://www.dlib.org/dlib/february00/vandesompel-oai/vandesompel-oai.html>> .
- TROLL, Denise; MOEN, Bill. Report to the DLF on the Z39.50 Implementers' Group. Acesso em: Disponível em: < <http://www.diglib.org/architectures/zig0012.htm>> . Acesso em: 22 ago. 2001.
- WARNER, Simeon. Exposing and harvesting metadata using the OAI Metadata Harvesting Protocol. *Tutorial, High Energy Physics Libraries Webzine*, v. 4, June 2001. Disponível em: < <http://library.cern.ch/HEPLW/4/papers3>> . Acesso em: 18 nov. 2001.
- WATERS, Donald J. The metadata harvesting initiative of the Mellon Foundation. *ARL Bimonthly Report*, n. 17, Aug. 2001. Disponível em: < www.arl.org/newsltr/waters.html> . Acesso em: 15 fev. 2002.
- WEIBEL, Stuart. Metadata: the foundations of resource description. *D-Lib Magazine*, July 1995. Disponível em: < <http://www.dlib.org/dlib/July95/07weibel.html>> . Acesso em: 5 maio 2000.
- ZILMAN, John. *Conhecimento público*. Belo Horizonte : Itatiaia/São Paulo : USP, 1979.