Bibliotecas digitais

Saberes e Práticas

Carlos H. Marcondes
Hélio Kuramoto
Lídia Brandão Toutain
Luís Sayão [orgs.]

ibict
EDUFBA

CARLOS H. MARCONDES

Professor do Departamento de Ciência da Informação da UFF e do Programa de Pós-Graduação em Ciência da Informação – PPGCI – IBICT/UFF. Mestre e Doutor em Ciência da Informação pela UFRJ/IBICT. Pesquisador do CNPq. Tem atuado em diferentes projetos relacionados a tópicos como bibliotecas digitais e documentos eletrônicos, como na Biblioteca Digital Brasileira de Teses e Dissertações (http://www.ibict.br/bdtd/acesso.htm); no Programa PROSSIGA; no projeto SciELO/Open Archives (http://www.scielo.br). Autor de várias publicações, em periódicos nacionais e internacionais, na área de bibliotecas digitais, metadados e publicações eletrônicas.

HÉLIO KURAMOTO

Formado em Engenharia Elétrica pela UnB em 1988. Diplomado em Estudos Aprofundados (DEA) em Ciências da Informação e da Comunicação pela ENSSIB (França), em 1995. Doutor em Ciências da Informação e da Comunicação pela Université Lumière (Lyon, França), 1999.
Trabalha no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) desde fevereiro de 1983, inicialmente como analista de sistemas e, em seguida, ocupou diversos cargos de confiança. Desde outubro de 2002, atua como Coordenador Geral de Projetos Especiais e Diretor substituto do IBICT.

Editora da UFBA
Rua Barão de Geremoabo,
s/n – Campus de Ondina
CEP 40170-290 – Salvador – BA
Tel: +55 71 3263-6164
Fax: +55 71 3263-6160
www.edufba.ufba.br
edufba@ufba.br

Instituto Brasileiro de Informação em
Ciência e Tecnologia - IBICT
SAS Quadra 05 Lote 06 Bloco H
70070-914 - Brasília, DF
Tel: +55 61 217-6360 / 6350
Fax: +55 61 217-6490
www.ibict.br
webmaster@ibict.br

Bibliotecas digitais

Saberes e Práticas

Carlos H. Marcondes
Hélio Kuramoto
Lídia Brandão Toutain
Luís Sayão [orgs.]

Salvador/Brasília
UFBA/IBICT
2005

Copidesque e Revisão (IBICT)
Francisco de Paula
Margaret de Palermo Silva

Revisão Editorial (EDUFBA)
Tânia de Aragão Bezerra
Magel Castilho de Carvalho

Capa
Joe Lopes

Projeto Gráfico e Editoração
Joe Lopes

Parcerias:
IBICT – Instituto Brasileiro de Informação em Ciência e Tecnologia, Programa de Pós-Graduação em Ciência da Informação/UFBA e a Editora da Universidade Federal da Bahia – EDUFBA

Biblioteca do IBICT

B582

Bibliotecas digitais: saberes e práticas/organizadores, Carlos H. Marcondes, Hélio Kuramoto, Lídia B. Toutain, Luís Sayão [prefácio de Aldo de Albuquerque Barreto]. – Salvador, BA : EDUFBA; Brasília: IBICT, 2005.
278 p: il.

Publicação em parceria da UFBA com o IBICT.
ISBN 85-232-0350-8

1. Biblioteca digital. 2. Biblioteca digital - Brasil. 3. Preservação digital. 4. Software livre. 5. Web semântica. 6. Metadados. 7. Comunicação científica - Tecnologia digital. I. Título. II. Marcondes, Carlos Henrique. III. Kuramoto, Hélio. IV. Toutain, Lídia Brandão. V. Sayão, Luís Fernando. VI. Barreto, Aldo de Albuquerque. VII. Universidade Federal da Bahia. VIII. Instituto Brasileiro de Informação em Ciência e Tecnologia.

CDU 027-021.131

Sumário

3. Dimensões do Uso

4. Gestão em Bibliotecas Digitais

5. Experiências Brasileiras e Internacionais

Prefácio

As mudanças na tecnologia da informação ocorridas durante os últimos anos reorganizaram as atividades associadas à ciência da informação. A sociedade sempre foi mais afetada pelas transformações da tecnologia do que pelo seu conteúdo, pelo menos no curto prazo. Por isso, aqueles que convivem mais de perto com estas alterações enfrentam com maior carga as conseqüências sociais e físicas de uma enorme ansiedade tecnológica.

O profissional desta área foi precipitado em uma conjunção de transformações, muitas das quais, ele ainda nem mesmo percebeu.

O modelo tecnológico inovador é fechado e induz a um distanciamento alienante de como ele opera ou se opera no melhor sentido. O conhecimento interno da técnica é, muitas vezes, considerado irrelevante e até indesejável. Se as suas conseqüências são benéficas para a sociedade, questionar é quase inoportuno.

A chegada da sociedade eletrônica de informação modificou drasticamente a delimitação de tempo e espaço da informação. A importância do instrumental da tecnologia da informação forneceu a infra-estrutura para modificações, sem retorno, das relações da informação com seus usuários.

A interatividade ou interatuação multitemporal mudou o acesso do usuário à informação para o tempo real, o que representa o tempo de acesso no entorno de zero nos estoques digitais de informação; possibilita o acesso em múltiplas formas de interação entre o receptor e a estrutura da informação contida neste espaço. A interatividade modifica o fluxo: usuário > tempo > informação. Esta condição reposiciona os acervos, o acesso e a distribuição da informação. O próprio documento de informação se torna mais acessível e libera o receptor das diversas funções em linha e em tempo linear, passando para um *acesso on-line* e com linguagens interativas; a interatividade em tempo real liberta o indivíduo dos seus rituais de sincronismo cotidiano: todos executando a mesma atividade e ao mesmo tempo: ir ao banco, ir ao trabalho, ir ao mercado, ir à aula.

A interconectividade reposiciona a relação usuário > espaço > informação, permitindo uma mudança estrutural no fluxo de informação que se torna multiorientado. Quando o tempo se aproxima de zero e a velocidade do infinito, os espaços se desterritorializam, perdem seus limites. A interconectividade dá ao indivíduo uma nova condição de contigüida-

de, onde a possibilidade de vizinhança se estende para a região do infinito e permite ao usuário da informação ter a possibilidade de deslocar-se, no momento de sua vontade, de um espaço de informação para outro espaço de informação; de um estoque de informação para um outro estoque de informação. O usuário passa a decidir *na escolha* de sua informação, é o determinador de suas *necessidades*. Ele agora é o julgador da relevância do documento que procura e da qualidade do estoque que o contém em tempo real, como se estivesse colocado virtualmente dentro do sistema de armazenamento e recuperação da informação.

Estas mudanças operadas no *status* tecnológico das atividades de armazenamento e transmissão da informação vêm trazendo mutações contínuas na relação da informação com seus usuários. Destacamos como instabilidades mais notáveis: a) mudanças na estrutura de informação; b) mudanças no fluxo de informação; c) o novo profissional da informação.

A interação em tempo real com a estrutura e suporte da informação tem questionado o caráter alfabético e linear do documento texto. O computador permite um texto livre das amarras da composição e da interpretação linear. O código lingüístico comum permanece como base das estruturas de informação, como o elemento sistemático e compulsório, para uma determinada comunidade de informação, mas os significados estão cada vez mais individualizados e intencionais. O documento em hipertexto permite que cada receptor modifique a mensagem arbitrariamente segundo seu conceito de relevância, atuando também como se fosse um autor do próprio texto. O fluxo da informação entre os estoques digitais e os receptores permeiam dois critérios: o da tecnologia da informação que almeja possibilitar o maior e melhor acesso disponível e o critério da ciência da informação, que intervém para qualificar este acesso em termos das competências individuais dos usuários para assimilar a informação.

Nas décadas iniciais da atividade, as unidades de informação trabalhavam com um fluxo de informação que era realizado em um tempo linear e direcionado a um único espaço de informação. Hoje, com a informação digital *on-line*, os fluxos de informação são multidirecionados, levando a meandros digitais onde os espaços são de vivência pela não presença.

O profissional desta área se encontra, na atualidade, como se em um ponto no presente entre o passado e o futuro. Convive com tarefas e técnicas tradicionais de sua profissão, mas precisa atravessar para uma

outra realidade, onde estão indo seus clientes, e aprender a conviver com o novo e o inusitado, numa constante renovação da novidade.

O presente livro trata de todos estes aspectos, no seu tema de bibliotecas digitais, desde a história destas bibliotecas, seus conceitos e definições, a articulação tecnológica, a gestão, o uso e as experiências brasileiras.

Certamente, um livro que preenche uma lacuna na área. A parte sobre DIMENSÃO CONTEXTUAL analisa o significado econômico, social e cultural da *Web* semântica e a recuperação de informações: ontologias, agentes, metadados e publicações digitais; os principais padrões de bibliotecas digitais. A DIMENSÃO TECNOLÓGICA verifica o que é necessário para montar o "site" de uma Biblioteca Virtual: os componentes da "URL" de uma biblioteca digital; rede, servidor, roteador, endereço IP, servidores de bancos de dados, motores de busca, a interface com o usuário. DIMENSÕES DO USO indica o novo papel das bibliotecas digitais na comunicação científica, suas finalidades, mecanismos, canais, filtros; o papel dos serviços de informação e das publicações eletrônicas; a auto publicação, o auto-arquivamento e informação livre; movimento "Open Access", "Open Archives". GESTÃO EM BIBLIOTECAS DIGITAIS explora os fluxos de trabalho no contexto de uma Biblioteca Digital: seleção, aquisição e inclusão de material digital; relações com os usuários; habilidades do profissional de informação no planejamento e operação de bibliotecas digitais. EXPERIÊNCIAS BRASILEIRAS e internacionais informa sobre as experiências de sucesso no Brasil e no exterior.

Aldo de Albuquerque Barreto
Professor, doutor e coordenador do Programa de Pós-Graduação do IBICT/UFF

Introdução

Propõe-se este livro a atender à necessidade existente na literatura de Ciência da Informação no Brasil sobre bibliotecas digitais. Por ser a primeira iniciativa neste sentido, inscreve-se como uma contribuição ao conhecimento do assunto. Assim, visa permitir a profissionais e estudantes da área uma compreensão abrangente de aspectos teóricos, técnicos e práticos envolvidos no planejamento e operação de bibliotecas digitais.

A partir de uma proposta inicial dos organizadores de agenda destas questões, culminando em um acordo de co-edição entre o Programa de Pós-Graduação em Ciência da Informação do ICI/UFBA e o IBICT, vários autores, especialistas no assunto, foram convidados a colaborar. A grande maioria respondeu positivamente, o que permitiu a abrangência do quadro de questões e problemas tratados nos diferentes capítulos.

O que primeiro nos parece relevante considerar diz respeito ao alcance e impacto das bibliotecas digitais na sociedade brasileira. A nosso ver, esse alcance ainda é bastante reduzido, chegando a ser mínima sua repercussão.

As bibliotecas sempre foram, historicamente, instituições que concentram a informação num lugar físico para servir a uma comunidade de usuários. Como as bibliotecas eram físicas, o alcance de seus serviços ficava restrito às comunidade que a elas conseguiam ter acesso.

Com o surgimento da Internet, esta situação evolui de forma drástica: não apenas o potencial de coletar e concentrar informações dispersas aumentou enormemente, como se tornou expressiva a capacidade de atender ao público no sentido mais amplo possível. As bibliotecas digitais tornam-se, desse modo, um instrumento poderoso de distribuição, cooperação e acesso ao conhecimento, atendendo e podendo servir de foco agregador a uma comunidade segmentada, distribuída geograficamente.

Na era global e informacional em que vivemos, torna-se imperiosa a necessidade de bibliotecas digitas num país como o Brasil, e nos mais distintos campos de atividades – a exemplo dos níveis fundamental e médio do ensino, dos serviços de saúde, do pequeno empreendedorismo, da produção cultual e da preservação do meio ambiente, para só citar alguns.

Que se atente para o fato de que a Internet não tem dono. A conseqüência daí resultante é a quantidade gigantesca de informação livre, disponível, de padrões e ferramentas de *softwares* também livres e gratui-

tos, o que é focalizado em alguns capítulos deste livro. Visto que se trata de informação perdida porque dispersa, desorganizada, avulta a importância da criação de bibliotecas digitais.

A consideração desses fatores, problemas e carências, ensejou o projeto de elaboração dos textos reunidos neste volume, na perspectiva do desenvolvimento de bibliotecas digitais no país.

Espera-se que este livro possa instigar e servir de ferramenta para fomentar discussões sobre o tema.

Carlos H. Marcondes *Hélio Kuramoto* *Lídia Brandão Toutain* *Luís F. Sayão*

1. Dimensão Contextual

Biblioteca digital: definição de termos

Lídia Maria Batista Brandão Toutain *

Este texto tem por objetivo auxiliar a compreensão e aquisição da terminologia própria da informática e da informação no que concerne ao conceito específico de biblioteca digital.

Organizado como um manual de consulta, este livro reúne a maior concentração de termos utilizados pelos autores dos textos presentes, além dos empregados por especialistas na área da biblioteca digital. Abrange também termos e conceitos nem sempre constantes de livros de textos ou de dicionários de informática.

É crucial que os usuários saibam em que consistem as diferenças entre aplicações da Web e Web e Web semântica, entendam a necessidade ou vantagens de utilizar certos padrões na implantação de uma *biblioteca digital, familiarizem-se com as novas tecnologias,* como os metadados, as linguagens documentais de disseminação, transmissão e uso.

As referências cruzadas ou *links* a outras entradas ou termos ajudarão a entender melhor determinado termo. Alguns termos têm mais de um significado, e cada qual é esclarecido para facilitar o entendimento.

O que aqui é apresentado resultou, principalmente, da contribuição valiosa dos outros autores deste livro.

Analógico

Sistema de representação de fenômenos por meio de analogias ou semelhanças.

---

* Doutora em filosofia, professora do Instituto de Ciência da Informação (ICI) da Universidade Federal da Bahia-UFBA. lbrandao@ufba.br

Assinatura digital

Modalidade de assinatura eletrônica, resultado de uma operação matemática que utiliza algoritmo de criptografia e permite aferir, com segurança, a origem e a integridade de um determinado documento digital.

Biblioteca digital

Biblioteca que tem como base informacional conteúdos em texto completo em formatos digitais – livros, periódicos, teses, imagens, vídeos e outros –, que estão armazenados e disponíveis para acesso, segundo processos padronizados, em servidores próprios ou distribuídos e acessados via rede de computadores em outras bibliotecas ou redes de bibliotecas da mesma natureza.

Certificado digital

Documento emitido e assinado digitalmente por uma autoridade certificadora que contém dados que identificam seu titular.

Criptografia

Codificação de dados segundo um código secreto, chamado chave, de forma que somente os usuários autorizados podem restabelecer sua forma original para consultá-lo.

Cookie

Arquivo que o navegador pode ler e registrar no computador do usuário um *website*. Assim, o usuário pode ser identificado, na segunda visita, no *website*. Corresponde a um arquivo de texto pequeno, que o servidor cria no disco rígido do usuário, sem sua permissão ou conhecimento, o que é freqüentemente chamado de "personalização".

Digital

Forma de codificação de objetos do mundo real por meios de dígitos binários – seqüência de 0´s e 1´s.

Digitalização

Processo de conversão de um documento analógico para um formato digital, convertendo-o em sinais binários, por meio de dispositivo apropriado, como um *scanner* ou câmera fotográfica digital.

Documento digital

Registro de informação codificado por meio de dígitos binários.

DOI – Digital Object Identifier

Sistema para identificar univocamente objetos digitais na Web, que tem como propósito principal a gestão da propriedade intelectual e o comércio digital dos objetos. Marca registrada da DOI Foundation (http://www.doi.org).

Dublin core

Padrão de metadados voltado para a descrição e descoberta de recursos na Internet. Possui um vocabulário controlado padronizado correspondente a 15 elementos de dados, que servem para descrever recursos web, como páginas HTML.

Formato de arquivo

Especificação de regras e padrões descritos formalmente para interpretação das cadeias de *bits* que formam os arquivos digitais. São classificados segundo o conhecimento e a propriedade dessas especificações: a) abertos – quando as especificações são de conhecimento público (exemplo: XML, HTML e TXT; b) fechados – quando as especificações não são divulgadas pelos proprietários (exemplo: DOC e PPT); c) proprietários – quando uma organização detém os direitos sobre o formato (exemplo: PDF e DOC). Quando um formato é produzido por um órgão de normalização, ele é classificado também como padronizado.

Frames/quadros

Técnica utilizada na organização de *websites* complexas (grandes empresas, jornais *on-line*) que consiste na divisão da tela em vários campos. A configuração mais popular é a repartição em um quadro fino na parte esquerda ou superior da tela (sumário com *links*) e um quadro maior para o anúncio dos arquivos endereçados por *links*.

Gestão de conteúdos

Um termo geral que se refere à organização, categorização e estruturação de recursos de informação – texto, imagem, som etc. – de

forma que possam ser armazenados, publicados, reutilizados, e editados com maior flexibilidade.

Hiperficção/hyper fiction

Utilização literária do hipertexto. Em vez de marcação linear, que permite ao autor uma planificação rigorosa do curso de recepção, são oferecidas ao leitor alternativas de ligação que exigem a renúncia momentânea, as ofertas específicas, e acentuam o processo de leitura, respectivamente, a situação, interação x imersão. O encontro com o texto é assim objetivo não só a respeito da geração do sentido, mas também em relação à composição do texto.

Hiperlink/hyperlink

Uma palavra, frase ou imagem que recebe uma marcação especial para funcionar como um elo com outro documento que pode estar no mesmo computador ou em outro servidor da Internet. O *hiperlink* é acionado por um clique do *mouse*.

Hipermídia/hypermedia

Ampliação do conceito de hipertexto. Forma de estruturação de documentos segundo o qual vários meios de armazenamento e transmissão de informação são integrados através de *hiperlinks*, permitindo a utilização simultânea de texto, sons, imagens e vídeo.

Hipertexto/hypertext

Forma de estruturação da informação que permite a leitura não linear de um texto, por meio de acionamento de *hiperlinks* que viabilizam a conexão direta com outras partes do documento ou com outros documentos disponíveis na Web.

HTML/Hypertext Markup Language

É a língua franca para publicação de documentos na Web. É um formato não-proprietário baseado no padrão SGML e pode ser criado e processado por uma grande variedade de ferramentas. O HTML utiliza *tags*, como <h1> e </h1>, para estruturar o texto em cabeçalhos, parágrafos, listas, *links* de hipertextos etc.

Icone/Ícon

Pictograma que, na tela, simboliza um objeto específico, programa ou funções, as quais são ativadas por clique, como,por exemplo, o símbolo da lixeira para a função rejeitar.

ICP – Infra-estrutura de Chaves Públicas

É um conjunto de técnicas, práticas e procedimentos, que estabelecem os fundamentos técnicos e metodológicos de um sistema digital baseado em certificação chaves públicas.

Internet

Rede das redes. Rede de computadores de abrangência mundial que interliga os mais diferentes sistemas computacionais e redes, e possibilita, por meio de protocolos padronizados, tais como o Transmission Control Protocol (TCP) e Internet Protocol (IP), a troca de dados entre eles. Os diferentes serviços de Internet baseiam-se nos próprios protocolos de transmissão: Simple Mail Transfer Protocol (SMTP) para e-mail, File Transfer Protocol (FTP) para transmissão de dados, Internet Relay Chat (IRC) para Internet, Chat e Hyper Text Transfer Protocol (HTTP) para World Wide Web.

Metadados

Elementos de descrição/definição/avaliação de recursos informacionais armazenados em sistemas computadorizados, organizado por padrões específicos, de forma estruturada.

Migração

Estratégia de preservação que consiste em copiar, converter e/ou transferir a informação digital de uma plataforma tecnológica que está se tornando obsoleta para uma outra mais atualizada e de uso corrente. O objetivo da migração é preservar a integridade de objetos digitais enquanto mantém a capacidade do usuário de recuperá-los, exibi-los e utilizá-los em face das constantes mudanças tecnológicas.

Multimídia

Tecnologia que permite ao computador trabalhar simultaneamente e de forma interativa com diversos tipos de registro informacional , como texto, som, imagens estáticas, animação e vídeo.

Navegadores/Web browsers

Programas de computadores que atuam como interface entre o usuário e a Web, permitindo-o navegar – ou surfar – pelos recursos oferecidos Web. No modelo cliente/servidor, o navegador se caracteriza funcionalmente como um programa cliente. Netscape, Mosaic, e Internet Explorer são exemplos de navegadores Web.

Objeto digital

No contexto dos arquivos e bibliotecas digitais, é um registro de informação codificado digitalmente, consistindo de conteúdo informacional, metadados e identificador.

Ontologia

1. Proposição evidente ou que se dá por verdadeira em um sistema lógico e da qual derivam dedutivamente outras proposições. Estabelece fundamentos de significados conceituais sem os quais a Web Semântica não seria possível.

2. Concepção de estruturas concebidas como um conjunto de relações entre elementos com funções definidas.

Open archives/arquivos abertos

Arquivo eletrônico baseado no protocolo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)–, geralmente de acesso livre, dotado de dispositivos de autopublicação e intercâmbio de metadados. Constitui um marco histórico do desenvolvimento da comunicação científica,da publicação eletrônica e das bibliotecas digitais. A partir dos Open Archives, estabeleceram-se padrão e protocolo para permitir interprobabilidade entre os sistemas das bibliotecas digitais.

Open source

Refere-se a pacote de *software* cuja distribuição acompanha o seu código fonte, o que possibilita ao usuário alterar e adequar o *software* segundo as suas necessidades.

OWL/Web Ontology Language

Padrão para elaborar ontologias na Web.Possui uma terminologia para formalizar a definição do conceito.

Padrões abertos

Referem-se a padrões cuja documentação está publicada e prontamente disponível, e que têm o desenvolvimento baseado em processos consensuais envolvendo as diversas partes interessadas.

Preservação digital

Conjunto de ações técnicas, gerenciais e administrativas destinadas a manter a integridade e a acessibilidade de objetos digitais de valor contínuo, pelo tempo que transcenda as mudanças tecnológicas.

RDF/Resource Description Framework

Uma infra-estrutura técnica desenvolvida pela W3C – baseada em XML– voltada para a descrição, intercâmbio de metadados e interoperabilidade. Tem importância fundamental na concepção da Web semântica.

Serviço de referencia digital

Serviço de intermediação entre os recursos informacionais disponíveis na Internet e os usuários. Tem como objetivo prestar assistência ao usuário no direcionamento de questões às fontes de informações Web e/ ou a outras fontes. Pode ser gerenciado automaticamente, através de *software* aplicativos ou através da assistência de especialistas.

Semântica

Disciplina que estuda a linguagem natural e formal (signos, termos, palavras) a sua função de representação, o que significa ou a que se refere.

Servidor web/web sever

Servidor que, posto à disposição dos computadores descentralizados (clientes), processa os serviços e envia o documento solicitado.

SGML/Standard Generalized Markup Language

Padrão internacional para definir descrições de estrutura e conteúdo de diversos tipos de documentos. Forma a base para o HTML e o XML.

Protocolo soap

Simple Object Access Protocol – SOAP – é um protocolo baseado em

XML usado para requisitar serviços Web, e intercambiar dados e informações estruturados em ambientes descentralizados e distribuídos.

Taxonomia

Ferramenta que possui a função de organização sistemática de conteúdos informacionais, apresentando as relações hierárquicas entre os conteúdos, classificando-os em grupos ou categorias.

Tesauro

Vocabulário controlado, compreendido como estruturas terminológicas, que visa a Padronizar a linguagem, em serviços de informação, cobrindo um domínio específico do conhecimento,traduzido de uma linguagem natural para uma linguagem de máquina.

Tesauro conceitual

Recurso que permite evidenciar melhor as relações entre os termos representados na ontologia*, privilegiando-se os aspectos do significado.

VRML/Virtual Reality Modeling Language

Linguagem que produz mundos tridimensionais e deixa-se representar por um correspondente *plug-in* também do *Browser*.

Unicode (ISO 10646-1 Universal Character Set)

Código de caracteres de 16 *bits* que busca cobrir todos os sistemas de escrita em escala mundial; deve substituir o ASCII – conjunto de caracteres de 7 bits que tem como limite somente 128 caracteres.

URI/Uniform Resource Identifier

Conjunto genérico de todos os nomes/endereços que identificam recursos informacionais na Web. Inclui URL´s e URN´s.

URN/Uniform Resource Name

Padrão da W3C que define uma sintaxe para nome e endereço de objetos digitais persistentes disponíveis na Internet.

URL – Uniform Resource Locator

Endereço eletrônico que especifica – em ordem: o protocolo de comunicação, o *host* servidor, o caminho do diretório, o nome do arquivo e o tipo do arquivo. Quando um desses elementos muda, o *link* é quebrado e se torna inoperável.

Virtual

Que não existe de fato. Representação eletrônica de algo real.

Web semântica/ semantic web

1. Evolução da WEB* atual, cujos proponentes foram Tim Berners-See, Hendler e Lassila. Visa fornecer estruturas e dar significado semântico ao conteúdo das páginas Web, criando um ambiente onde agente de *software* e usuários possam trabalhar de forma cooperativa.

2. Conhecimento semântico* estruturado.

Webmaster

Pessoa responsável pela manutenção técnica, observação e atualização de um *Website*, sem que seja necessariamente o autor.

W3C

World Wide Web Consortium é um consórcio internacional de indústrias que desenvolve padrões e protocolos comuns que promovem a evolução da WWW e assegura a sua interoperabilidade.

W3

Uma rede de servidores conectados por meio de um protocolo comum, permitindo acesso a milhões de recursos informacionais hipertextuais. Também conhecida como WWW, Web e *World Wide Web*.

XML/Extensible Markup Language

1. Linguagem de padrão universal, referendada pela *w3c* e aberta, que descreve documentos eletrônicos nos quais o conteúdo e sua descrição compõem um único arquivo.

2. Estrutura sintática de padrão que descreve dados entre aplicações e recursos de máquina. Constitui a base de vários padrões na área de informação.

3. Linguagem genérica que descreve estrutura de documentos eletrônicos, padrão simples, em formato texto. Possui recursos bem definidos e extensíveis que permitem descrever e pesquisar objetos, atributos e valores através do relacionamento entre eles.

Referências

CÂMARA TÉCNICA DO DOCUMENTO ELETRÔNICO. *Glossários de Documentos Arquivísticos Digitais*. Rio de Janeiro: CONARQ, julho 2004. Disponível em: <http://www.arquivonacional.gov.br/conarq/cam_tec_doc_ele/download/Glossario_CTDE_2004.pdf>. Acesso em: 01/06/2005.

COSTA, Sely. Mudanças no processo de comunicação científica. In: Muller, S.; Passos, E. *Comunicação científica*. Brasília, 2000.

WATERS, D. What are Digital Libraries? *CLIR ISSUES*, Washington, DC, n.4, jul/aug ., 1998. Disponível em *www**.clir. org/ pubs/issues/. (Acesso em 2005).

TRISKA, R.; CAFÉ, L. Arquivos abertos: subprojeto da Biblioteca Digital Brasileira. CI. Inf., Brasília, v.30, n.3, set/dez. 2000.

ARMS, W. Key concepts in the architecture of the digital library. *D-LIB* MAGAZINE. Virginia, USA. Disponível em http://*www*.dlib . org/dlib/july95/07 arms. (Acesso em 2005).

BULLOCK, A. *Preservation of digital information*. Ottawa: National Library of Canada, apr. 22, 1999. http://*www*.nic -ben.ca/publication/pl-259.html.

PFAFFENBERGER, B. *Webster's new world: dicionário de informática*. 6. ed. Rio de Janeiro Campus, 1998.

MARCONDES, C.; SAYÃO, L. Integração e interabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. *Ci. Inf.*, Brasília, v.30, n.3, p. 24-33, 2001.

MARDERO, M. Serviço de referência virtual. *Ci. Inf.*, Brasília, v.30, n.2, p.7-15,2001.

Estado atual das bibliotecas digitais no Brasil

*Murilo Bastos da Cunha, Ph. D.**
*Cavan McCarthy, Ph. D.***

1. Introdução

A história da Internet no Brasil está intimamente ligada à implantação da Rede Nacional de Pesquisa (RNP) [URL: http://www.rnp.br], cuja criação remonta a 1989 (Histórico, 2003). Muitos estados criaram suas próprias redes, enquanto a RNP era responsável pela interligação dos sistemas estaduais e também pelo acesso ao exterior.

Entre 1991 e 1993, a RNP implantou o seu *backbone* básico, interligando 11 estados. Neste período, a rede era utilizada somente para comunicação científica e tecnológica. O segundo período, de 1994 a 1996, foi marcado por uma enorme expansão de suas atividades. Em maio de 1995, a Internet brasileira foi aberta para uso comercial, provocando um enorme crescimento da demanda que, até aquele momento, estava reprimida. A Internet, a partir de então, foi rapidamente adotada no Brasil como uma nova modalidade de comunicação, completamente apropriada para os avanços tecnológicos e para uma nação recentemente democratizada (Cunha, 1999).

*Universidade de Brasília, Departamento de Ciência da Informação e Documentação. Brasília DF 70910-900 Brasil. murilobc@unb.br.
**Louisiana State University, School of Library and Information Science 267 Coates Hall, Baton Rouge, LA 70803-3920, USA. mccarthy@lsu.edu.

As taxas de crescimento da Internet brasileira foram consideravelmente superiores às ocorridas em outros países (Histórico, 2003). Em 2000, cerca de 10 milhões de brasileiros utilizavam a Internet (Pastore, 2000 a). Nesse mesmo ano, o Brasil atingia cerca de 41% dos usuários latino-americanos da Rede (Pastore, 2000b). De acordo com dados do Ibope (2002), o número de cidadãos com acesso à Internet em residências alcançava 14 milhões, dos quais 7,8 milhões poderiam ser considerados internautas ativos, com um tempo de navegação que já ultrapassava dez horas por mês.

Entretanto, apesar do crescimento da Internet no Brasil ser acentuado, esse dado estatístico pode ser enganador, pois para a imensa maioria da população, o acesso ao computador ainda é um sonho. Os dados do IBGE (IBGE, PNAD), referentes a 2002, atestam que dos 170 milhões de habitantes do país, apenas 3,3% (o equivalente a quase 7 milhões de habitantes) dispunham de computadores em casa. Este é um percentual pequeno se comparado com a existência no domicílio de telefone (fixo ou celular) e aparelho de televisão, respectivamente em 62 e 90, em cada 100 residências no país. Além disso, o acesso à Internet e computadores, isto é, da chamada inclusão digital, é irregular no que concerne às classes sociais. Em 7 de outubro de 2003, segundo a Agência Nacional de Telecomunicações (ANATEL), do total de internautas brasileiros, 42% pertenciam à classe A, que tem 5% da população; a classe B, que representa 19% dos brasileiros, representava 48,7% dos internautas e, finalmente, as classes C, D e E, onde vivem 76% da população, contavam com apenas 9,3% dos internautas. O diagnóstico da ANATEL para a pequena quantidade de internautas no Brasil em relação à população é de que o problema em aumentar o acesso não está na telefonia, mas nos preços dos microcomputadores.

Em setembro de 2004, segundo Daniele Madureira, a Internet já estava presente em 15% dos lares brasileiros, mas

> em um País no qual a desigualdade social chega a ser pungente e apenas um terço dos 182 milhões de habitantes tem condições reais de adquirir o básico para a subsistência, vale a pena investir em uma mídia que só atinge, segundo pesquisa da Fundação Getúlio Vargas (FGV), 15% das residências no Brasil?

Continuando, a autora comenta que

> (...) no mundo, foi preciso 38 anos para o rádio atingir 50 milhões de pessoas, a TV aberta levou 13 anos e a TV a cabo, 10. Para atingir a mesma marca, a Internet precisou de apenas cinco anos. (...) lembrando que ao considerar o acesso à internet independentemente da origem (casa, trabalho, escola, *cyber café*) o número de usuários no Brasil é muito maior do que 27 milhões. Destaca que a rede não se limita aos PCs: a mídia inclui celulares, *palms*, videogames, TV digital e sistemas pervasivos (encontrados em carros e aparelhos eletrodomésticos). No nosso País, o número reduzido de internautas é um reflexo da exclusão social e econômica de grande parte da população, mas estão sendo adotadas políticas para reverter isso. (Madureira, 2004)

Na mesma época (setembro de 2004), escrevendo sobre o comércio eletrônico no Brasil, Dubes Sonego, aponta os crescimentos verificados no número de usuários residenciais e nas diversas formas de acesso à Rede. Esses dados constam da Tabela 1.

Tabela 1 – Usuários ativos da Internet

Período	Usuários ativos (milhões)	Banda larga (milhões; % do total de usuários)	Linha discada (milhões; % do total de usuários).
Maio de 2002	7,59	1,5 (19,76 %)	6,09 (80,24 %)
Maio de 2003	7,96	2,11 (26,6 %)	5,84 (73,4 %)
Maio de 2004	11,68	4,93 (42,2 %)	6,75 (57,8 %)

Fonte: Sonego (2004)

Os dados estatísticos da Tabela 1 demonstram que o número de usuários da banda larga cresceu 228,23 % entre maio de 2002 e maio de 2004. Além disso, esses dados numéricos sinalizam que as perspectivas para expansão são excelentes. Uma análise da Internet brasileira preparada por Howton, Wardwell & Gunday (2001) mostrou que o papel do Estado na provisão de incentivos para a indústria de informática poderá fazer com que o país possa ter uma revolução na Internet. Para que os projetos de bibliotecas digitais alcancem sucesso junto à comunidade brasileira, é vital que haja uma diminuição da chamada exclusão digital, fazendo com que se tenha uma ampla capilaridade da Rede em todos os segmentos sociais.

2. Desenvolvimento da biblioteca digital no Brasil

O Brasil possui uma tradição de serviços bibliotecários, funcionando na maioria das cidades de médio e grande porte; que, geralmente, contam com sistemas de automação de bibliotecas (McCarthy, 1990; McCarthy; Schmidt, 1994; Corte, A. e outros, 2003). As bibliotecas suportam os programas educacionais, especialmente os de segundo e terceiro graus.

Na última década, as bibliotecas digitais tiveram um significativo impacto no setor de biblioteca e informação, notadamente na América do Norte, onde atraíram enorme atenção (Chowdhur; Chowdhury, 1999). O rápido avanço da Internet no Brasil, conforme já apontado no item anterior, e existência de uma base razoável de bibliotecas automatizadas, naturalmente irá redundar na ampliação do número de bibliotecas digitais. No contexto atual, as maiores iniciativas brasileiras se enquadram em quatro categorias: ciência e tecnologia, educação, literatura e humanidades, história e política. A seguir, serão analisados os principais projetos institucionais de bibliotecas digitais.

Ciência e Tecnologia (C&T)

IBICT

O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), uma agência do Ministério da Ciência e Tecnologia (MCT), sempre desempenhou um papel vital no desenvolvimento de atividades informacionais avançadas. Os seus projetos que chegaram à fase de implementação serão comentados a seguir.

O Programa de Informação e Comunicação para a Pesquisa (PROSSIGA) [URL: http://prossiga.ibict.br] foi criado em 1995, no âmbito do Ministério da Ciência e Tecnologia; em meados de 2001, foi transferido para o IBICT. Ele é um portal que tem por objetivos a divulgação da informação, comunicação e inovação para a ciência e tecnologia. Além de manter um diretório com ponteiros para sítios selecionados nas diversas áreas de C&T, possui também uma série de bibliotecas digitais [http://www.prossiga.br/bvtematicas/], denominadas "bibliotecas virtuais". Muitas dessas bibliotecas são guias de sítios Web sobre cada um dos temas e que, geralmente, incluem dados sobre: pesquisadores, associações e sociedades científicas, instituições de ensino, publicações, legislação, principais periódicos e obras de referência. Essas bibliotecas, na

verdade, são diretórios de sítios Web relacionados com um tema específico, geralmente incluindo dados sobre: pesquisadores, associações e sociedades científicas, instituições de ensino, publicações, legislação, principais periódicos e obras de referência. Elas cobrem uma variedade de assuntos e que foram criadas contando, em sua maioria, com a cooperação de importantes instituições.

Em maio de 2003, o Prossiga tinha uma média diária de 85.980 acessos (Prossiga em números, 2003).

Até janeiro de 2005, eram 19 os temas cobertos pelas bibliotecas digitais. Abaixo, em ordem alfabética, são informados o tema, a data de criação, o URL e resumo do projeto.

1) Agropecuária na Amazônia (28 de março de 2003) [URL: http://www.prossiga.br/embrapa/agropecuaria/]: foi organizada pelo Centro de Pesquisa da Amazônia Oriental da EMBRAPA. Ela disponibiliza informações sobre ciências agrárias e setor produtivo, com foco no agronegócio.

2) Astronomia (14 de agosto de 2001) [URL: http://www.prossiga.br/astronomia/]:, projeto do Observatório do Valongo da Universidade Federal do Rio de Janeiro, em parceria com o IBICT. O seu objetivo é disponibilizar informações sobre a astronomia, não só para o pesquisador, mas também para o usuário leigo.

3) Ciências Sociais (17 de novembro de 1999) [URL: http://binac.nce.ufrj.br/cienciassociais/]: projeto do Instituto de Ciências Sociais da Universidade Federal do Rio de Janeiro, que disponibiliza informações em sociologia, antropologia e ciência política.

4) Economia (13 de janeiro de 1998) [URL: http://www.prossiga.br/nuca-ie-ufrj/economia/]: projeto do Instituto de Economia da Universidade Federal do Rio de Janeiro, visa disponibilizar informações sobre os diferentes domínios da ciência econômica.

5) Educação (2 de setembro de 1998) [URL: http://bve.cibec.inep.gov.br]: projeto do Instituto Nacional de Pesquisas Educacionais Anísio Teixeira (INEP), que disponibiliza informações sobre avaliação de ensino; estatísticas educacionais; ensino fundamental, médio e superior; outros assuntos correlatos à Educação.

6) Educação a Distância (13 de maio de 1998) [URL: http://

www.prossiga.br/edistancia/]: projeto da Faculdade de Educação e Instituto de Ciência da Informação da Universidade Federal da Bahia; disponibiliza informações sobre recursos tecnológicos da comunicação e informação na educação, especialmente aqueles relacionados com televisão, vídeo, computador e conexão à Internet.

7) Energia [URL: http://www.prossiga.br/cnencin/bvenergia/]: foi desenvolvida pelo Centro de Informações Nucleares da Comissão Nacional de Energia Nuclear (CIN/CNEN). Ela inclui "sítios que tratem de área de energia nos aspectos tecnológicos da produção, conversão e utilização eficiente de todas as fontes de energia, convencionais e não convencionais, incluindo impactos ambientais, econômicos e sociais" (Comissão Nacional de Energia Nuclear. CIN. Biblioteca Virtual de Energia. O que é?)

8) Engenharia Biomédica (6 de julho de 2000) [URL: http://www.prossiga.br/fem-unicamp/bvbiomedica/]: projeto da Faculdade de Engenharia Mecânica da UNICAMP; disponibiliza informações nas áreas de bioengenharia, biomateriais, próteses e orteses, biomecânica e assuntos correlatos à engenharia biomédica.

9) Engenharia do Petróleo (13 de maio de 1998) [URL: http://www.prossiga.br/dep-fem-unicamp/petroleo/]: projeto do Departamento de Engenharia do Petróleo da Universidade Estadual de Campinas; disponibiliza informações sobre os variados aspectos da engenharia do petróleo, tanto nacional quanto internacional.

10) Estudos Culturais (3 de julho de 2000) [URL: http://binac.nce.ufrj.br/estudos culturais/]: projeto conjunto do Programa Avançado de Cultura Contemporânea e do Programa de Pós-Graduação da Escola de Comunicação da Universidade Federal do Rio de Janeiro. Os assuntos cobertos são: gênero e sexualidade, identidades nacionais, pós-colonialismo, etnia, cultura popular, pós-modernidade, multiculturalismo e globalização.

11) Inovação Tecnológica (15 de dezembro de 2000) [URL: http://www.prossiga.br/finep/]: projeto conjunto entre a Financiadora de Estudos e Projetos (FINEP) e o Prossiga; disponibiliza informações sobre inovação tecnológica em seus múltiplos aspectos.

12) Jurídica (1° de dezembro de 1998) [URL: http://www.cjf.gov.br/bvirtual/]: projeto do Conselho da Justiça Federal; disponibiliza o aces-

so direto a documentos (textos doutrinários, periódicos eletrônicos, códigos legais) e a fontes secundárias (bibliotecas, referência de jurisprudência, legislação, andamento processual e instituições jurídicas).
13) Literatura (27 de maio de 2002) [URL: http://binac.nce.ufrj.br/literatura/]: parceria entre o Programa Avançado de Cultura Contemporânea da Universidade Federal do Rio de Janeiro e o Programa de Pós-Graduação em Teatro da UNIRIO. Ela disponibiliza informações sobre literatura, com ênfase na brasileira, latino-americana e portuguesa.
14) Mulher (13 de setembro de 2001) [URL: http://www.prossiga.br/bvmulher/cedim/]: projeto do Conselho Estadual dos Direitos da Mulher (CEDIM/RJ); disponibiliza informações sobre as mulheres relacionadas com: saúde, violência, trabalho, cultura, direito e cidadania, educação, poder e participação política.
15) Museus de Ciência e Divulgação Científica (20 de agosto de 2001) [URL: http://binac.nce.ufrj.br/cienciassociais/]: parceria entre a Fundação Oswaldo Cruz e o IBICT; disponibiliza informações sobre museus das áreas de ciências naturais e físicas.
16) Óptica Básica e Aplicada (24 de abril de 1998) [URL: http://www.prossiga.br/ifsc-usp/optica/]: projeto do Instituto de Física de São Carlos, órgão da Universidade de São Paulo; disponibiliza informações sobre o apoio o desenvolvimento do ensino e pesquisa na área de Óptica.
17) Política Científica e Tecnológica (dezembro de 1996) [URL: http://www.prossiga.br/politica-ct/]: lançada em pelo próprio Prossiga.
18) Saúde Mental (20 de agosto de 2001) [URL: http://www.prossiga.br/ee_usp/saudemental/]: projeto conjunto do Departamento de Enfermagem Materno Infantil e Psiquiatria da Escola de Enfermagem da Universidade de São Paulo e do Ministério da Saúde; disponibiliza informações sobre a promoção, prevenção e tratamentos referentes ao melhoramento, manutenção ou a restauração da saúde mental.
19) Saúde Reprodutiva (7 de junho de 1999) [URL: http://www.prossiga.br/fsp_usp/saudereprodutiva]: projeto da Faculdade de Saúde Pública da Universidade de São Paulo; disponibiliza informações sobre os eventos biológicos, psicológicos e sociais relacionados à reprodução humana.

Outra importante iniciativa dentro do Prossiga é a série de bibliotecas digitais [http://www.ibict.br/notaveis/] dedicadas a importantes cientistas brasileiros. Nessas bibliotecas, geralmente são incluídos dados biográficos, produção intelectual do pesquisador (em sua maioria com textos completos), matérias na imprensa, imagens, fotos e informações sobre suas descobertas e/ou contribuições para a ciência e tecnologia. Até janeiro de 2005 as sete bibliotecas existentes eram:

a) Anísio Teixeira [URL: http://www.prossiga.br/anisioteixeira/]: foi lançada em 10 de dezembro de 1997, pelo Prossiga com a colaboração de 14 entidades que cederam documentos, além de arquivos pessoais privados. Ela disponibiliza referências, cartas e documentos do educador Anísio Teixeira, incluindo sua produção intelectual, correspondência, matérias e entrevistas na mídia, depoimentos e homenagens ao educador (instituições que receberam o seu nome, prêmios instituídos e recebidos pelo pensador).
b) Carlos Chagas [URL: http://www.prossiga.br/chagas/]: é um projeto realizado pela Fundação Oswaldo Cruz (RJ) e que disponibiliza informações sobre o médico Carlos Chagas e sobre a doença descoberta pelo mesmo (a Doença de Chagas).
c) Gilberto Freyre [URL: http://prossiga.bvgf.fgf.org.br/portugues/index.htm]: foi organizada pela Fundação Gilberto Freyre (Recife, PE), com o objetivo de servir de portal oficial de disseminação do arquivo documental da Fundação. Além da documentação impressa, são interessantes as coleções especiais com fotos, áudio e vídeos de/sobre o pensador brasileiro.
d) Leite Lopes [URL: http://www.prossiga.br/leitelopes/]: foi lançada em 24 de março de 1999, pelo Prossiga e pelo Centro Brasileiro de Pesquisas Físicas (CBPF). Ela disponibiliza informações sobre o físico José Leite Lopes, relacionadas com os seguintes aspectos: trajetória, contribuição à Física, produção científica (geral, sobre Física e sobre política científica), presença na mídia, textos biográficos e pintura.
e) Oswaldo Cruz [URL: http://www.prossiga.br/oswaldocruz/]: foi lançada em 19 de outubro de 2000, pela Fundação Oswaldo Cruz (RJ). Ela disponibiliza informações sobre o médico Oswaldo Cruz e seu trabalho em prol da saúde pública no Brasil.

f) Otto Gottlieb [URL: http://www.prossiga.br/ottogottlieb/]: foi lançada em 27 de maio de 2003, pelo Prossiga com o apoio da FAPERJ. Ela disponibiliza informações sobre o professor de química Otto Gottlieb, estando organizada em seis partes: trajetória, produção científica (1945/2002), legado científico (teses e dissertações orientadas), presença na mídia e depoimentos de colaboradores, amigos e ex-alunos.

g) Paulo Freire [URL: http://www.paulofreire.ufpb.br/paulofreire/index.html] projeto conjunto, desenvolvido em 2000, pelas áreas de Educação, Informática, Educação a Distância da Universidade Federal da Paraíba, em cooperação com o CNPq. O seu objetivo é disponibilizar os pressupostos filosóficos, sociológicos e pedagógicos do pensamento desse importante educador.

h) Vital Brazil [URL: http://www.prossiga.br/vitalbrazil/]: foi lançada em 4 de julho de 2002, pelo Prossiga em cooperação com o Instituto Vital Brazil e a Fundação Oswaldo Cruz. Ela disponibiliza informações sobre o cientista Vital Brazil, incluindo sua produção intelectual, biografia e sobre ofidismo e soro antiofídico.

Biblioteca Digital Brasileira – No final de 2001, o IBICT lançou o portal Biblioteca Digital Brasileira (BDB) [http://www.ibict.br/bdb/portal/bdb_portal.htm], que oferece uma interface para acesso a teses e dissertações digitalizadas. A BDB pretende permitir o acesso, num único lugar, a artigos científicos, teses, dissertações, além dos arquivos digitais dos órgãos da área de C&T do Governo; incluirá também salas de bate-papo e listas de discussão para sociedades científicas.

Biblioteca Digital de Teses e Dissertações – A implantação da BDTD remonta a 1995 quando o IBICT integrou, numa única base de dados, as referências bibliográficas de teses e dissertações de 17 universidades brasileiras. Em dezembro de 2002, o IBICT instalou a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) [URL: http://bdtd.ibict.br/bdtd], cuja missão é viabilizar a criação de um consórcio de publicações eletrônicas com a finalidade de localizar e disponibilizar teses e dissertações produzidas no Brasil.

Esse consórcio, inédito no Brasil, considera que potenciais provedores desse tipo de informação, em grande número, podem trabalhar de forma conjunta, proporcionando a multiplicação de pontos de acesso para

o usuário, com significativo aproveitamento dos recursos materiais e humanos disponíveis.

Além de reunir e integrar grande número de publicações eletrônicas de teses e dissertações, o consórcio permite também aos usuários o acesso aos textos integrais dos artigos do seu interesse, por meio do portal Biblioteca Digital Brasileira. Outra conseqüência positiva da criação desse consórcio é a incorporação das informações de teses e dissertações produzidas no Brasil ao sistema internacional Networked Digital Library of Theses and Dissertations.

Para acelerar a criação de bibliotecas digitais nos *campi* brasileiros, o IBICT desenvolveu o Sistema de Publicações Eletrônicas de Teses e Dissertações (TEDE). Este pacote de programas permite a rápida implantação desse tipo de atividade dentro da instituição de ensino superior, que fará o papel de coletora e provedora de dados. O IBICT receberá os dados coletados e fará a sua agregação, em nível nacional.

Vale ressaltar que a CAPES aprovou, no final de 2003 (Brasil iniciou...), a recomendação para que o depósito das teses e dissertações da BDTD seja obrigatório. Isto fará com que seja aumentado o acervo da biblioteca e que haja um maior interesse em iniciar projetos de teses digitais nas universidades brasileiras. Em 18 de janeiro de 2005, existiam 6.228 teses/dissertações em texto completo, hospedadas em 14 repositórios de instituições de ensino superior.

Arquivos Abertos – O IBICT implantou, no final de novembro de 2001, o projeto dos Arquivos Abertos que visa oferecer à comunidade científica uma alternativa de comunicação de textos acadêmicos, científicos e tecnológicos via ambiente web. O Arquivo Aberto do IBICT, denominado Diálogo Científico (DICI) [URL: http://dici.ibict.br/], utiliza o *software* eprints [URL: http://www.eprints.org/] desenvolvido pela University of Southampton (UK), cujos arquivos estão de acordo com os padrões do Metadata Harvesting. No sítio da eprints.org é possível visualizar as principais informações sobre os projetos desenvolvidos no Brasil.

O conceito da Iniciativa de Arquivos Abertos (Open Archives Initiative, OAI) representa o anseio da comunidade científica em formar um fórum aberto para aprimorar o desempenho do atual modelo de comunicação científica. Ele visa também formar um repositório de informações que disponibilize na Web, de forma pública e gratuita, as contribuições sub-

metidas diretamente pelos autores. Pretende-se, assim, acelerar o processo da divulgação dos resultados das pesquisas.

Até início de 2005, segundo Márdero Arellano (2005), o IBICT já tinha implantando o DICI nas seguintes entidades: a) Sociedade Brasileira de Genética [URL: http://sbg.ibict.br/]; b) Arena Científica (Arquivos Abertos para a Comunidade de Especialistas em Comunicação), em parceria com a Rede de Informação em Comunicação dos Países de Língua Portuguesa (PORTCOM), mantida pela Sociedade Brasileira de Estudos Interdisciplinares da Comunicação (INTERCOM) [URL: http://comunicacao.ibict.br/]; c) Diálogo Científico em Ciência da Informação [URL: http://dici.ibict.br/] para os pesquisadores brasileiros dessa área.

Repositório Institucional Digital do IBICT. Utilizando a plataforma DSpace, desenvolvida pela MIT, o IBICT implantou, no final de 2004, o Repositório Institucional Digital do IBICT (RIDI) [URL: https://dspace.ibict.br/dspace/]. Em 8 de fevereiro de 2005, o seu acervo contava com 13 documentos digitais.

SCIELO

Criada em 1965, com a denominação de Biblioteca Regional de Medicina (BIREME), ocupando as instalações da antiga Escola Paulista de Medicina (atualmente Universidade Federal de São Paulo). Em 1967, foi assinado um convênio entre o Governo Brasileiro e a Organização Pan-americana de Saúde (OPAS) e a BIREME expandiu sua abrangência geográfica, passando a funcionar como um centro internacional para o continente, como núcleo central de uma rede cooperativa que englobou as instituições de informação da área de saúde. Sua denominação passou a ser Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde [URL: http://www.bireme.br/], mantendo-se a sigla anterior.

A BIREME, localizada em São Paulo, é conhecida pela distribuição da informação biomédica em CD-ROM (Castro; Packer; Castro, 1989; McCarthy, 1996). Ela também tem sido um ativo centro da UNESCO, distribuindo os gerenciadores de bases de dados bibliográficos CDS/ISIS e MICROISIS.

A Biblioteca Científica Eletrônica *On-Line*, mais conhecida pela sigla em inglês SCIELO (Scientific Electronic Library On-line) [URL: http://www.scielo.org], é um projeto piloto desenvolvido em 1997 pela BIREME com o apoio financeiro da Fundação de Amparo à Pesquisa de São Paulo (FAPESP). Seu objetivo era desenvolver uma metodologia de digitalização,

armazenamento, disseminação e avaliação da literatura científica em formato eletrônico, por meio da provisão de uma biblioteca digital de periódicos eletrônicos com texto completo (Wladimir, 2002). O projeto piloto foi testado no período de março de 1997 a maio de 1998, com periódicos, em sua maioria das áreas de saúde; a partir de junho de 1998, passou a operar de forma regular, incorporando novos títulos de outras áreas. O sistema atualmente também inclui títulos em ciências sociais e humanidades. A coleção básica, em janeiro de 2005, era formada por 131 títulos brasileiros, 47 chilenos, 17 cubanos e 21 espanhóis. Os artigos em texto completo estão disponíveis nas versões em html e pdf, possuindo resumos em inglês. Desde 2002, o *Web of Science* faz hipervínculos com os artigos armazenados no SCIELO (ISI, 2002).

Segundo o levantamento estatístico mais recente, durante 2003, o SCIELO recebeu 2.788.451 acessos que possibilitaram o *downloading* de 42.372.055 páginas de artigos de periódicos (Statistics of www.scielo.br).

A coleção está sendo expandida para incluir periódicos da Costa Rica, Espanha e Venezuela. Um subsistema separado cobre periódicos de saúde pública. Uma característica interessante do SCIELO é a ligação entre o nome do autor do artigo e seu *curriculum vitae* armazenado no Lattes (Santana; Packer; Barretto, 2001), mantido pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico na base de dados denominada Plataforma Lattes [http://www.cnpq.br/]. Essa característica permite que o usuário do SCIELO possa encontrar no Lattes produtos de pesquisa do mesmo autor. A Scientific Development Network observou que o modelo SCIELO é uma alternativa viável para o aumento da visibilidade e do fator de impacto dos periódicos de países em desenvolvimento na comunidade internacional (Dickson, 2002).

Observatório Nacional

O Observatório Nacional criou a Biblioteca Digital de Obras Raras e Especiais [URL: http://www.docvirt.no-ip.com/obnacional/acervo.html], incluindo documentos relevantes sobre astronomia e história da ciência no Brasil.

Agência Nacional de Águas

A Agência Nacional de Águas lançou a Biblioteca Digital Proágua Semi-Árido [URL: http://www.ana.gov.br/proagua/biblioteca.asp] onde estão incluídos os textos completos de documentos e relatórios técnicos, bem como a legislação sobre água no Brasil.

Educação

Ministério da Educação

Portal da CAPES. A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [URL: http://www.capes.gov.br] é o órgão do Ministério da Educação responsável pela avaliação dos programas brasileiros de pós-graduação. Sua primeira iniciativa na área da informação digital foi a criação, em novembro de 2000, do Portal da CAPES [URL: http://www.periodicos.capes.gov.br] que oferece acesso a diversos sistemas europeus e norte-americanos de periódicos eletrônicos.

Em 2001, com um estoque de cerca de 1500 títulos, o portal tinha um uso médio de 350.000 acessos por mês; no final de 2002, passou a oferecer acesso a 2.400 títulos, dos fornecedores Science Direct, Blackwell, Ovid, ACM e outras fontes; em janeiro de 2005, passou a oferecer acesso a 8.466 títulos de periódicos nacionais e estrangeiros.

O Portal é um programa que tem financiamento federal e que permitia, em janeiro de 2005, o acesso a alunos universitários, professores e pesquisadores de 130 instituições, entre elas universidades, centros de pesquisa da EMBRAPA e centros federais de ensino técnico. Essas entidades podem consultá-lo por meio de um endereço Internet autorizado (IP address). Mediante pagamento, a CAPES passou a permitir o acesso às instituições privadas de ensino, sendo a Universidade Católica de Brasília a primeira organização a assinar este tipo de contrato.

No segundo semestre de 2003, com os problemas orçamentários advindos dos cortes feitos pelo Executivo, o Portal da CAPES passou por período difícil na renovação das assinaturas com os provedores de periódicos eletrônicos. Chegou-se a propalar até que o Portal estava sendo "fechado", havendo manifestações a favor do funcionamento do Portal por parte das sociedades, associações científicas e acadêmicos. Uma Comissão Consultiva para Negociação do Portal de Periódicos foi criada e, em 3 de outubro de 2003, informou que

> as negociações junto aos fornecedores apontam convergências para novos termos de contrato, em bases favoráveis ao País. A renovação dos atuais contratos, cujo valor anual é de 18 milhões de dólares, se articula com a possibilidade de inclusão de novos títulos, atendendo às demandas das áreas do conhecimento ainda não contempladas no Portal. (CAPES. Informativo n. 16)

No final de novembro de 2003, foi divulgada a notícia de que a renovação dos contratos foi aprovada (CAPES. Informativo n. 20).

Em janeiro de 2005, o Portal também provia acesso a 80 bases de dados bibliográficos, destacando-se: *Web of Science, Agricola, Biological Abstracts, MEDLINE, Psychological Abstracts, CAB Abstracts, COMPENDEX, Cambridge Scientific Abstracts e INSPEC.*

Base de dados sobre teses – Como parte de suas atividades na área de pós-graduação, a CAPES, ao longo dos anos, acumulou uma base de dados sobre teses e dissertações. Essa base de dados disponibilizada em linha em junho de 2002 [URL: http://www.capes.gov.br/Scripts/index.idc?pagina=/servicos/indicadores/TesesDissertacoes.htm], contém 185.000 registros de teses defendidas no Brasil a partir de 1987. É uma base de dados que não inclui acesso ao texto completo, informando, porém, o endereço das bibliotecas depositárias onde o documento está disponível.

Em meados de 2004, essa base de dados passou a oferecer acesso ao texto completo de teses e dissertações da área de História [URL: http://www2.liphis.com/]. Este é um projeto coordenado pela Universidade Federal do Rio de Janeiro que, no final de janeiro de 2005, já contava com 249 registros de diversas universidades.

Portal Domínio Público – Em novembro de 2004, o Ministério da Educação lançou o Portal Domínio Público [URL: http://www.mec.gov.br/dominiopublico/], que inclui uma biblioteca digital que reunirá mais de mil obras literárias, artísticas e científicas da cultura universal (Macedo, 2004). O acervo estará à disposição da população para pesquisa, estudo e compartilhamento de informações. As obras oferecidas pelo Portal já são de domínio público ou têm autorização legal de divulgação e exibição por parte dos detentores de direitos autorais ou representantes legais.

O acervo do Portal pretende incluir uma enorme variedade de documentos: romances, crônicas, poesia, livros e publicações de diversas áreas, teses e dissertações, fotografias, mapas, gravações fonográficas, filmes, programas de rádio e televisão. Como se pode notar pelos componentes projetados para o seu acervo, o Portal de Domínio Público tem metas de difícil atendimento, pois, mesmo nos países desenvolvidos, são poucos os exemplos de bibliotecas digitais que fizeram a integração de tantos tipos de suportes informacionais. Em janeiro de 2005, este projeto só tinha implementado um catálogo coletivo de obras

digitalizadas, disponíveis no Brasil e no exterior e que, mediante solicitação do usuário, poderiam ser feitos os *downloads* para armazenamento local. Espera-se que o Ministério da Educação consiga prover recursos financeiros, e principalmente humanos e tecnológicos, para atingir essas metas originais. Em caso contrário, o Portal pode ser transformar num mero catálogo coletivo de referências de documentos digitais.

Universidade de São Paulo

A Universidade de São Paulo (USP), a maior do país, com cerca de 40.000 estudantes (USP hoje, 2005), mantém o portal denominado Saber [URL: http://www.saber.usp.br/]. Por meio desse Portal, é possível acessar a coleção de teses e dissertações; obras raras e especiais; a Biblioteca Virtual do Estudante de Língua Portuguesa e o catálogo da Editora da USP (EDUSP), a maior editora universitária do Brasil, que publica cerca de 70 títulos por ano.

O Portal também oferece acesso ao Dedalus, o catálogo em linha do Sistema de Bibliotecas da USP (SIBI), contendo cerca de 1.400.000 registros bibliográficos. O SIBI é pioneiro na utilização do Aleph/Exlibris, programa de automação de bibliotecas, atualmente adotado por diversas instituições brasileiras.

Biblioteca Virtual do Estudante de Língua Portuguesa – Em 1989, a Universidade de São Paulo criou um centro educacional tecnologicamente avançado, a chamada Escola do Futuro. Em 1997, essa Escola disponibilizou na Internet a Biblioteca Virtual do Estudante Brasileiro (BibVirt) [URL: http://www.bibvirt.futuro.usp.br/], atualmente denominada Biblioteca Virtual do Estudante de Língua Portuguesa.

A BibVirt oferece cerca de 200 textos completos de autores clássicos das literaturas brasileira e portuguesa, nos formatos rtf e html. Textos que podem apoiar programas de televisão de ensino a distância também estão sendo digitalizados. Considerando a natureza colorida da cultura brasileira, fotografias de aves, frutas e música também podem ser acessadas. Além disso, existem arquivos de áudio com discursos de políticos e cientistas brasileiros.

Essa biblioteca tem ganhado prêmios, sendo votada nos concursos do I-Best como um dos mais belos sítios de sua categoria. Estatísticas recentes mostram que a maioria dos usuários é formada por estudantes e professores do primeiro grau das grandes cidades brasileiras; 7% dos seus usuários provêm de cidades com menos de 10.000 habitantes (Pro-

jeto, 2002), sendo, portanto, um interessante indicador da penetração da Web no interior brasileiro.

Biblioteca Digital de Obras Raras e Especiais – Em novembro de 2003, a Biblioteca Central da USP inaugurou a sua Biblioteca Digital de Obras Raras e Especiais [URL: http://www.obrasraras.usp.br/] (Moreno, 2003). Em janeiro de 2005, estavam disponibilizadas 41 obras em várias áreas do conhecimento, obedecendo aos critérios de antiguidade, valor histórico e inexistência de novas impressões ou edições do título. Os demais documentos do acervo, num total de 1.224 títulos, tiveram as capas digitalizadas.

Biblioteca de Teses e Dissertações – Um dos componentes do Portal do Saber, a Biblioteca Digital de Teses e Dissertações da USP [URL: http://www.teses.usp.br] disponibiliza o acesso ao texto completo a este tipo de documento armazenado em formato pdf (MASIERO, 2001).

Biblioteca Digital de Direitos Humanos – A Biblioteca Virtual de Direitos Humanos da Universidade de São Paulo [URL: http://www.direitoshumanos.usp.br/], criada pela sua *Comissão de Direitos Humanos*, coloca à disposição dos interessados as informações sobre este importante e atual tema.

Os documentos disponibilizados são: textos aprovados pelos organismos internacionais e ratificados pelo Governo brasileiro; textos e legislações elaboradas pelo sistema Interamericano de Direitos Humanos; texto completo de documentos importantes de interesse geral do cidadão brasileiro, tais como: Estatuto do Idoso, Estatuto da Criança e Adolescente, Código de Defesa do Consumidor.

Universidade Estadual de Campinas

A Universidade Estadual de Campinas (UNICAMP), um grande centro de pesquisa que contava, em 2003, com cerca de 30.000 alunos e cuja maioria está na área de pós-graduação (UNICAMP. Retrato).

Ela mantém um dinâmico sistema de bibliotecas [URL: http://www.unicamp.br/bc/] que, em 8 de novembro de 2001, ativou uma biblioteca digital, denominada Biblioteca Digital da Unicamp [URL: http://libdigi.unicamp.br], que tem por objetivo preservar e disseminar, "através da produção científica/acadêmica da Unicamp em formato eletrônico de: artigos, fotografias, ilustrações, obras de arte, revistas, registros sonoros, teses, vídeos e outros documentos de interesse ao desenvolvimento científico, tecnológico e sócio-cultural" (Apresentação da biblioteca). Além desses documentos, também estão sendo incluídos os

trabalhos apresentados em eventos e os periódicos eletrônicos editados pela universidade.

Foi utilizado um conjunto de programas desenvolvido localmente e denominado Nou-rau – um trocadilho do termo em inglês *know-how*. Nesse conjunto [URL: http://www.rau-tu.unicamp.br/nou-rau/sbu/], que segue a política do *software* livre, podem ser inseridos diversos tipos de documentos, de teses a recortes de jornais, em diferentes formatos, incluindo html, pdf, rtf e mesmo Powerpoint (Descrição, 2002).

Em janeiro de 2005, o acervo dessa biblioteca digital incluía 3.933 teses e 10.461 documentos gerais, tendo sido realizados 382.848 *downloads* desde a sua implantação.

Universidade Federal do Paraná

A Universidade Federal do Paraná utiliza a plataforma *DSpace*, desenvolvida pelo *Massachusetts Institute of Technology* (MIT) e Hewlett-Packard, para a criação de arquivos digitais em nível de universidades. O *DSpace* viabiliza a coleção, armazenagem, indexação, conservação e disponibilização da produção cientifica de professores e pesquisadores. O projeto, denominado Biblioteca Digital de Teses e Dissertações [URL: http://dspace.c3sl.ufpr.br/dspace/index.jsp], inclui teses, dissertações, relatórios técnicos, arquivos com sons e imagens. Em 25 de janeiro de 2005, o acervo total consistia de: 15 arquivos com sons e imagens; 3 relatórios técnicos; 53 teses e dissertações.

Universidade Federal do Rio Grande do Sul

A Universidade Federal do Rio Grande do Sul, com a colaboração do IBICT, lançou a sua Biblioteca Digital de Teses e Dissertações [URL: http://www.biblioteca.ufrgs.br/bibliotecadigital/], contendo documentos defendidos a partir de 2001. Em 25 de janeiro de 2005, estavam disponibilizados 2.420 documentos em texto completo.

Pontifícia Universidade Católica (Rio de Janeiro)

A Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), utilizando a metodologia da NLTD, lançou a sua Biblioteca Digital de Teses e Dissertações [URL: http://www.maxwell.lambda.ele.puc-rio.br/cgi-bin/db2www/PRG_0490.D2W/INPUT?CdLinPrg=pt]. Vale ressaltar que, a partir de agosto de 2002, todas as teses e dissertações defendidas na PUC-Rio obrigatoriamente passaram a ser depositadas nesse repositório.

Pontifícia Universidade Católica de Minas Gerais

A Pontifícia Universidade Católica de Minas Gerais (PUC-MG) lan-

çou, com a colaboração do IBICT, a sua Biblioteca Digital de Publicações [URL: http:// www2.sistemas.pucminas.br/BDP/SilverStream/Pages/pg_BDPPrincipal.html]. Nessa biblioteca, estarão disponibilizadas a teses e dissertações produzidas nessa universidade.

Universidade Federal de Santa Catarina

A Universidade Federal de Santa Catarina criou o seu Banco de Teses e Dissertações [URL: http://teses.eps.ufsc.br/index.asp]. A origem desse Banco está ligada à criação, em 1995, do Laboratório de Mídia e Conhecimento, vinculado ao Programa de Pós-Graduação em Engenharia da Produção. No período de 1995 a 1999, foram digitalizadas 283 dissertações de mestrado e 74 teses de doutorado.Em janeiro de 2005, esse repositório contava com 4827 itens digitais, sendo 4.005 teses e dissertações.

Universidade Estadual Paulista

A Universidade Estadual Paulista (UNESP) lançou a sua Biblioteca Digital, que utiliza o Sistema Nou-Rau desenvolvido pela Universidade de Campinas. Essa Biblioteca [URL: http://www.biblioteca.unesp.br/bibliotecadigital/] tem por objetivo o armazenamento, controle e disseminação de documentos digitais. O seu acervo é composto de teses e dissertações, partituras musicais, slides, fotografias, periódicos eletrônicos e a produção científica da instituição.

Universidade Federal de São Carlos – A Universidade Federal de São Carlos criou, em 17 de agosto de 2004, no seu Portal do Conhecimento, a Biblioteca Digital de Teses e Dissertações [URL: http://www.bdtd.ufscar.br/]. Essa biblioteca visa divulgar, em texto completo, a teses e dissertações defendidas nos seus 18 programas de pós-graduação.

Universidade do Vale do Rio dos Sinos – A Universidade do Vale do Rio dos Sinos (UNISINOS) criou, com o apoio do IBICT, a sua Biblioteca Digital de Teses e Dissertações [URL: http://bdtd.unisinos.br].

Universidade Católica de Brasília – A Universidade Católica de Brasília UCB) lançou, com o apoio do IBICT, a sua Biblioteca Digital de Teses e Dissertações [URL: http://www.bdtd.ucb.br/tede/tde_busca/index.php]. Em 2 de fevereiro de 2005, o seu acervo era composto de 40 teses de mestrado, em texto completo.

Universidade Católica de Pelotas – A Universidade Católica de Pelotas (Rio Grande do Sul) lançou, com o apoio do IBICT, a sua Biblioteca Digital de Teses e Dissertações [URL: http://200.17.170.152/tede/tde_busca/index.php].

Universidade Federal Fluminense – A Universidade Federal Fluminense, por meio do seu Núcleo de Documentação, lançou a sua Biblioteca Digital de Teses e Dissertações [URL: http://www.bdtd.ndc.uff.br]. Esse projeto contou com o apoio do IBICT.

Universidade Livre da Mata Atlântica – A Universidade Livre da Mata Atlântica, instituição de utilidade pública, sediada no Convento dos Franciscanos, em Cairú (Bahia), é uma entidade do terceiro setor, dedicada à promoção do desenvolvimento sócio-econômico-ecológico integrado. Contando com o apoio do Worldwatch Institute, sediado em Washington (USA), montou a Biblioteca Digital Wwi-UMA [URL: http://www.wwiuma.org.br/]. O acervo dessa biblioteca é constituído de textos completos de entrevistas, artigos de diversas publicações periódicas e os números publicados na língua portuguesa da revista *World Watch*.

Universidade Católica Dom Bosco – A Universidade Católica Dom Bosco, por meio do seu Mestrado em Desenvolvimento Local, criou, em maio de 2002, a sua Biblioteca Digital [URL: http://www.bdmdl.ucdb.br/]. No projeto foi utilizado o protótipo desenvolvido na Universidade Federal de Santa Catarina pelo grupo de pesquisa Bibliotecas Digitais da Rede Metropolitana de Alta Velocidade/Florianópolis. Até 25 de janeiro de 2005, existiam 395 documentos digitais no acervo dessa biblioteca.

Universidade do Vale do Itajaí – A Universidade do Vale do Itajaí, por meio do seu Centro de Ciências Tecnológicas da Terra e do Mar, criou a sua Biblioteca Digital [URL: http://www.cttmar.univali.br/~bibcttmar/]. O projeto foi desenvolvido por alunos de Ciência da Computação. O seu acervo inclui trabalhos de conclusão do curso de graduação, teses e dissertações.

Faculdade Integrada do Recife – A Faculdade Integrada do Recife criou a Biblioteca Digital da Copesq [URL: http://docente.fir.br/copesq/nipe/bibliotecadigital/] que inclui as publicações e monografias de conclusão de cursos dos alunos de graduação dessa instituição de ensino superior.

Instituto Tecnológico de Aeronáutica – O Instituto Tecnológico de Aeronáutica (ITA) lançou a sua Biblioteca Digital [URL: http://www.bibl.ita.br.], um projeto desenvolvido pela Biblioteca Central e a Divisão de Ciência da Computação. Em janeiro de 2005, constavam do seu acervo 97 trabalhos de conclusão de curso de graduação, 389 teses e 110 dissertações.

Literatura e humanidades

Biblioteca Nacional

A Biblioteca Nacional do Rio de Janeiro (BN) [URL: http://www.bn.br/], o maior acervo da América Latina (com cerca de nove milhões de peças), foi fundada em 1808 a partir da coleção da Biblioteca Real de Portugal que havia sido transferida de Lisboa para o Rio de Janeiro, por ocasião da invasão francesa à Península Ibérica.

Na BN é possível consultar incunábulos raros, apreciar as gravuras de Debret ou ouvir músicas compostas por *Villa-Lobos. Todos esses acessos são facilitados por meio do* Programa Biblioteca Nacional sem Fronteiras [URL: http://www.bn.br/fbn/bibsemfronteiras/], criado em janeiro de 2001, e que visa a montagem de uma biblioteca digital concebida de forma ampla como um ambiente onde estão integradas as coleções digitalizadas essenciais para o entendimento das culturas brasileira e portuguesa (Programa Biblioteca).

Periódicos raros do início do século XIX, gravuras e imagens também estão sendo digitalizados e indexados. Para visualizar essas imagens e realizar pesquisa textual é necessário fazer os *downloads* dos *plug-in* do programas DocReader , produzido pelo DocPro [URL: http://www.docpro.com.br/default. htm].

Na Biblioteca Nacional existe uma área separada denominada "Literatura Eletrônica", na qual os leitores que desejarem consultar textos completos de 15 dos principais autores da literatura brasileira, podem encontrá-los no formato pdf.

Em janeiro de 2005, já estavam digitalizados os *Anais da Biblioteca Nacional*, relativos aos volumes 1-117 e cobrindo o período de 1876 a 1997. Nessa mesma data, o catálogo da BN disponibilizava os seguintes tipos de documentos digitalizados: 167 livros, 504 manuscritos, 196 fotos, 17 partituras musicais, 10 mapas antigos, 10 materiais audiovisuais, seis obras raras (de um acervo total de 42.000 itens).

Portais gerais e de humanidades

Os portais dos grandes provedores brasileiros de acesso à Internet oferecem serviços relevantes à literatura e humanidades. O Terra/Lycos [URL: http://www.terralycos.com], megaportal com sede em Barcelona (Espanha), é considerado o portal líder nos mercados espanhol e brasileiro. Ele reivindica contar com mais de 400 milhões de visitantes em mais de 40 países. No Brasil, ele

mantém um dinâmico portal [URL: http://www.terra.com.br] que oferece uma livraria eletrônica, a Virtual Books Online [URL: http://virtualsbooks.terra.com.br/], que também contém uma coleção com 120 obras clássicas das literaturas portuguesa e brasileira, além de ficção, poesia e livros infantis.

O outro grande portal brasileiro é o Universo *On-line* (UOL) [URL: www.uol.com.br/] que divulga ser o mais utilizado sítio na portuguesa, com mais de um milhão de assinantes, sete milhões de páginas e 10 milhões de usuários únicos visitantes a cada mês. O UOL, criado em 28 de abril de 1996, é uma empresa mantida pelo grupo Abril (editor, entre outros, da revista *Veja*) e *Folha de S. Paulo*. Grande parte do conteúdo do UOL é restrito aos assinantes; o acervo é composto, entre outros, de obras de referência, dicionários eletrônicos, revistas e jornais.

Outros recursos em humanidades

Pequenos sítios, sem apoio financeiro oficial, disponibilizam textos completos da literatura brasileira que estão em domínio público. É bom lembrar que a legislação brasileira sobre direitos autorais permite a cópia livre de conteúdos após 70 anos do falecimento do autor, além disso, os clássicos da literatura brasileira geralmente fazem parte da bibliografia básica dos vestibulares para ingresso nas instituições de ensino superior. Os principais sítios são:

a) Biblioteca Virtual de Literatura [URL: http://www.biblio.com.br]: criada em setembro de 2000, provê acesso a textos em formato html, com biografias e comentários dos principais literatos brasileiros e portugueses, com trechos de partes de suas obras que já estão em domínio público. Ela tem uma freqüência média diária de 350 acessos.
b) Alexandria Virtual [URL: http://www.supervirtual.com.br]: a Editora Supervirtual é uma empresa privada de armazenamento digital, localizada em São Paulo (SP). Foi além dos clássicos básicos, disponibilizando uma ampla variedade de material em diferentes formatos (zip, doc e pdf). Em janeiro de 2005, o seu acervo era de 102.396 itens.
c) EBooks Brasil [URL: http://www.ebooksbrasil.com/] o maior distribuidor de livros eletrônicos no país.
d) Biblioteca Digital Pausa para a Filosofia [URL: http://www.bdpfilosofia.pop.com.br/]: provê acesso ao texto completo de livros em português, dos principais filósofos estrangeiros.

História, Direito e Política

Senado Federal

O Senado Federal, com sede em Brasília, tem uma longa tradição na área de informática. O Centro Gráfico do Senado Federal [URL: http://www.senado.gov.br/web/cegraf/conselho/] oferece acesso no formato pdf a textos completos de livros que cobrem cerca de 500 anos do Brasil, relacionados com as áreas de história, política e direito.

O Senado Federal mantém o Sistema de Informações do Congresso Nacional (SICON) [URL: http://www6.senado.gov.br/sicon/] que permite o acesso a quatro bases de dados (Bibliotecas da Rede Rubi, Discursos dos Senadores, Legislação federal, Matérias com tramitação no Senado). Essas bases de dados são integradas e muitos dos seus registros encaminham para o texto completo do documento, por exemplo: discurso do parlamentar, texto de uma norma legal ou artigo de periódico.

Fundação Getúlio Vargas

A Fundação Getúlio Vargas mantém no Rio de Janeiro o Centro de Pesquisa e Documentação de História Contemporânea do Brasil (CPDOC) [URL: http://www.cpdoc.fgv.br/comum/htm/] que focaliza a recente história política brasileira.

Os usuários podem consultar materiais relacionados com a história brasileira mediante registro gratuito. Eles têm acesso a cerca de 35.000 fotografias digitalizadas, áudio com 400 entrevistas de personalidades que participaram de eventos políticos nos últimos 50 anos, outros materiais audiovisuais e reproduções de documentos impressos. Também é possível consultar um catálogo de quase um milhão de manuscritos e verbetes do *Dicionário histórico-biográfico brasileiro*.

Superior Tribunal de Justiça

A Biblioteca Digital Jurídica [URL: http://bdjur.stj.gov.br/dspace/index.jsp] foi criada em 22 de setembro de 2004, no âmbito do Superior Tribunal de Justiça. Ela utiliza a Plataforma DSpace, desenvolvida no MIT. Em 8 de fevereiro de 2005, constavam em seu acervo 42 documentos digitais em texto integral, composto de normas legais, artigos de periódicos, palestras, discursos, conferências e capítulos de livros.

Artes e Arquitetura

Instituto Moreira Sales

O Instituto Moreira Salles (Rio de Janeiro e São Paulo) [URL: http://www.ims.com.br] iniciou, em fevereiro de 2004, a digitalização dos acervos da sua divisão musical, que tem sob sua guarda as coleções dos músicos Antônio D'áuria, Pixinguinha, Elizeth Cardoso e Ernesto Nazareth, do crítico José Ramos Tinhorão, do pesquisador Humberto Franceschi e do jornalista Walter Silva (Conde, 2002). O projeto visa disponibilizar para o público um riquíssimo acervo composto de livros, partituras e gravações musicais raras que contam a história do choro, do surgimento das escolas de samba, do nascimento do rádio e dos antigos carnavais do Rio de Janeiro.

Tesouros da Cidade de São Paulo

Em 11 de março de 2003, foi lançado o Projeto Tesouros da Cidade de São Paulo: Biblioteca Digital Multimídia [URL: http://portal.prefeitura.sp.gov.br/secretarias/cultura/bibliotecas/marioandrade/0009]. Esse Projeto, uma parceria entre a Secretaria Municipal da Cultura de São Paulo e o Instituto Embratel 21, visa a digitalização e a veiculação pela Internet de acervos raros localizados na Biblioteca Mário de Andrade e no Departamento de Patrimônio Histórico.

Os acervos incluem cerca de 5.000 gravuras e fotos das transformações urbanas ocorridas em São Paulo, no período de 1860 a 1960 e os textos integrais de 120 livros raros sobre o Brasil, publicados entre 1551 a 1885.

Biblioteca Virtual do Amazonas

A Biblioteca Virtual do Amazonas [URL: http://www.bv.am.gov.br/portal/] foi lançada em abril de 2002. Trata-se de um projeto da Secretaria de Cultura do Estado do Amazonas que tem por objetivo disponibilizar parte do acervo de materiais raros da Biblioteca Pública do Estado e do Instituto Geográfico e Histórico do Amazonas. Abrange textos em domínio público sobre temas amazônicos, documentos da Província (1852-1888), estudos literários e recortes de jornais. Em janeiro de 2005, já estavam disponibilizados 174 imagens de projetos arquitetônicos de edificações de Manaus no final do século 19 e início do século 20, além de plantas da cidade, linhas de bonde, relatórios, mensagens, exposições e legislação da Província do Amazonas.

3. Conclusões

Existe uma série de fatores que favorecem o desenvolvimento da biblioteca digital no Brasil. Conforme já exposto no primeiro item, o país está formalmente comprometido com o amplo acesso à Internet; cerca de 20% da população – em torno de 35 milhões de pessoas – deveria ter acesso no final de 2004.

As empresas telefônicas, desde 2000, passaram a oferecer acesso em banda larga e serviços de ADSL a partir de 2002. O Ministério da Ciência e Tecnologia está implantando a Rede Nacional de Pesquisa –2 (RNP-2), o equivalente à Internet-II (Histórico, 2001), que interligará todos os 27 estados brasileiros com ligações até 155 mps, com conexões para as redes estrangeiras.

Apesar de recente, a biblioteca digital começa a receber apoio financeiro governamental. Algumas bibliotecas digitais que obtiveram sucesso foram criadas baseadas nas experiências e recursos existentes em suas instituições, por exemplo: Prossiga, Biblioteca Virtual do Estudante de Língua Portuguesa e SCIELO. Em alguns casos, esse sucesso faz com que a biblioteca digital se transforme no núcleo central de um portal, por exemplo: o SCIELO. Esforços cooperativos também desempenharam importante papel na criação de bibliotecas digitais no Brasil, ressaltando-se o caso do Prossiga.

Existem claras similaridades entre as bibliotecas digitais no Brasil e em outros países, são elas: o suporte governamental tem sido vital na implantação adequada dos sistemas; instituições nacionais e as bibliotecas nacionais têm estado em primeiro plano no movimento das bibliotecas digitais. A própria Biblioteca Digital de Saúde pode ser comparada ao US PubMed [URL: http://www4.ncbi.nlm.nih.gov/PubMed], o Hospital Virtual [URL: http://www.vh.org] ao UK OMNI [URL: http://omni.ac.uk]. Em muitos países, a Biblioteca Nacional ocupa um papel de liderança na construção da rede de bibliotecas digitais. Espera-se que a Biblioteca Nacional do Rio de Janeiro assuma um papel relevante no Brasil.

É claro que existem limitações ao desenvolvimento das bibliotecas digitais no Brasil. Uma diferença entre a América do Norte é que, nos Estados Unidos, o acesso a fotografias históricas é um grande elemento incentivador das bibliotecas digitais. Muitas das coleções do American Memory da Library of Congress [URL: http://memory.loc.gov] baseiam-se

em coleções de fotografias históricas. Uma possível causa desse fato poderia ser que as fotografias antigas teriam mais dificuldades de serem preservadas nas condições climáticas brasileiras. Outra razão adicional poderia ser o fato de que, no passado, os jornais brasileiros talvez tenham operado com relativamente baixo nível técnico e, portanto, menores chances para acumularem grandes coleções fotográficas.

E as perspectivas futuras? A maioria das estatísticas demonstra que as bibliotecas digitais brasileiras estão tendo significativos impactos na sociedade, na pesquisa e cultura do subcontinente. O atual sistema de bibliotecas digitais parece cobrir de forma razoável as áreas prioritárias, principalmente pelo fato de que foi criado por instituições que já estavam atendendo as demandas existentes.

Ainda não está claro como as bibliotecas digitais podem se desenvolver sem uma coordenação adicional. Além disso, também não está claro se essa coordenação deva ser espontânea ou estimulada centralmente. Até meados de 2004, uma variedade de instituições estava disseminando suas teses e dissertações, tanto no formato de resumo como em texto completo. Com o advento da Biblioteca Digital de Teses e Dissertações (BDTD) o IBICT recuperou o tempo perdido, fazendo com que o país possa ser mais bem servido por meio de um sistema coordenado para controle e acesso desse tipo de material. É o caso de indagar se ações similares deveriam ser copiadas em outros segmentos, como o feito pelo Governo Federal ao aprovar, em 2002, normas gerais para a criação de sítios governamentais (Brasil, 2002).

Em fevereiro de 2005, época em que este capítulo foi escrito, o Governo do Presidente Lula já tinha ultrapassado o seu segundo ano de administração e a temática da inclusão digital ou a ampliação do acesso ao mundo virtual está sendo retomado pelas autoridades federais. Infelizmente, o Programa Sociedade da Informação (Governo, 2002), lançado pelo governo anterior (Fernando Henrique Cardoso), parece que ficou "congelado" e as verbas oriundas do Fundo de Universalização dos Serviços de Telecomunicações (FUST) não foram gastas até agora. É sabido que nesses recursos seriam contemplados projetos para dar acesso à Internet às bibliotecas públicas, escolas e telecentros. Assim, recursos financeiros que poderiam estimular o surgimento de projetos de bibliotecas digitais oriundos de bibliotecas públicas estão no compasso de es-

pera, aguardando mudanças por parte dos ministros envolvidos. Além disso, recursos orçamentários têm sido contingenciados.

De qualquer modo, apesar da presente situação financeira, à semelhança de outros países, é visível que a Internet e as bibliotecas digitais poderão executar um importante papel na garantia da ampliação do acesso à informação no Brasil. A informatização do país e a criação de telecentros somente terão sentido se for possível oferecer materiais informativos de boa qualidade à população. As bibliotecas digitais constituem o único canal que tem o potencial de disponibilizar conteúdo cultural de bom nível para a Nação em geral. Isso representaria a grande contribuição das bibliotecas digitais para o progresso brasileiro.

Referências

ANATEL. Só 8% dos brasileiros têm acesso à internet. Brasília, 7 de out. de 2003. <URL: http://www.anatel.gov.br> Acesso em: 23 de dezembro de 2003.

Apresentação da Biblioteca Digital da Unicamp. <URL: http://www.unicamp.br/bc/bibdig/apresentacao.htm> Acesso em: 21 de janeiro de 2005.

BRASIL. Presidência da República. Gabinete Civil. Comitê Executivo do Governo Eletrônico. Resolução E-Gov n. 7, de 30 de julho de 2002, estabelece regras e diretrizes para os sítios na Internet da Administração Pública Federal. Disponível em: <http://www.governoeletronico.e.gov.br/governoeletronico/publicacao/down_anexo.wsp?tmp.arquivo=E15_223resolucao_n7_sitios_oficiais2562002.pdf > Acesso em: 11 de fevereiro de 2005.

CAPES. Assessoria de Comunicação. *Informe CAPES, n. 16*. Brasília, 10 de outubro de 2003.

CAPES. Assessoria de Comunicação. *Informe CAPES, n. 20*. Brasília, 28 de novembro de 2003.

CAPES. Assessoria de Comunicação. *Informe CAPES n. 4*. Brasília, 18 de março de 2004.

"Capes vai ampliar Portal de periódicos". *Jornal da Ciência: JC E-Mail*, n. 1929, 2001. Disponível em:http://200.177.98.79/jcemail/anteriores/jcemail2001/jcdiario-06-12-01.html#18 Acesso em: 10 de abril de 2003).

CASTRO, R. C. F., PACKER, A. L., CASTRO, E. Projeto LILACS/CD-ROM - Literatura Latino-Americana e do Caribe em Ciências da Saúde em disco compacto. *Revista Brasileira de Biblioteconomia e Documentação*, v. 22, n. 1/2, p. 105-114, 1989.

COMISSÃO Nacional de Energia Nuclear. Centro de Informações Nucleares. Biblioteca Virtual de Energia.<URL: http://www.prossiga.br/cnencin/bvenergia/> Acessado em: 26 de março de 2004.

CHOWDHURY, G. G.; CHOWDHURY, S. Digital library research: major issues and trends". *Journal of Documentation*, v. 55, n. 4, p. 409-448, 1999.

CONDE, Miguel. A história da música brasileira a um clique. *Globo Online, Rio de Janeiro, 19 de fevereiro de 2004*.

CORTE, A. R. *Avaliação de software para bibliotecas e arquivos: uma visão do cenário nacional*. 2ª. ed. São Paulo: Polis, 2002. 221 p.

CUNHA, M. B. "Desafios na construção de uma biblioteca digital". *Ciência da Informação*, v. 28 n. 3, p. 255-266, 1999. Disponível em: <http://www.ibict.br/cienciadainformacao/include/getdoc.php?id=631&article=323&mode=pdf> Acesso em: 11 de fevereiro de 2005.

Descrição do Nou-Rau. Disponível em: <http://www.rau-tu.unicamp.br/nou-rau/desc-pt.html> Acesso em: 11 de fevereiro de 2003.

DICKSON, David. Thumbs up for electronic publication. SciDev.Net. 2002 Feb. 4. Disponível em: <http://www.scidev.net/Editorials/index.cfm? fuseaction=readEditorials&itemid=23&language=1> Acesso em: 11 de fevereiro de 2005.

Governo Lula promete ampliar programas para inclusão digital. Último Segundo, 27 dez. 2002. Disponível em: <http://ultimosegundo.ig.com.br/useg/economia/mundovirtual/artigo/0,,1038174,00.html> Acesso em: 11 de fevereiro de 2005.

"História da RNP." Disponível em: <http://www.rnp.br/rnp/historico.html> Acesso em: 11 de fevereiro de 2005.

HOWTON, N.; WARDWELL, J.; GUNDAY, H. "Brazilian Internet". Disponível em: <http://www.stanford.edu/class/las194/GroupProjects01/brazil.pdf> Acesso em: 11 de fevereiro de 2005.

IBGE. PNAD 2002. <URL: http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2002/default.shtm> Acesso em: 11 de fevereiro de 2005.

Ibope. "Internet residencial: pela primeira vez brasileiros navegam mais de dez horas por mês". Disponível em: <http://www.ibope.com.br/calandraWeb/servletCalandraRedirect?temp=5&proj=PortalIBOPE&pub=T&db=caldb&comp=Not%EDcias&docid=84F9B44A9B934071832 56ECA00657A73> Acesso em: 1 de fevereiro de 2005.

"ISI Web of Science inaugura links para os artigos da SciELO". *SciELO News*, n. 4, 2002.

MACEDO, Lilian. MEC lança portal Domínio Público com mais de mil obras digitalizadas. Agencia Brasil / Radiobrás. 2004 nov. 16. <Disponivel em: <http://www.radiobras.gov.br/materia_i_2004.php?materia=207381&editoria=> Acesso em: 11 de fevereiro de 2005.

MADUREIRA, Daniele do Nascimento. Uma mídia presente em 15% dos lares. *Meio & Mensagem*, 13 de setembro de 2004.

MÁRDERO ARELLANO, Miguel Ángel. Mensagem enviada em 25 de janeiro de 2005.

MASIERO, P.C. e outros. A Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo. *Ciência da Informação*, v. 30, n. 3, p. 34-41, 2001. Disponível em:<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=225&layout=abstract> Acesso em: 28 de janeiro de 2005.

MCCARTHY, C. Brazilian CD-ROMs as informational and reference sources. IN: Seminar on the Acquisition of Latin American Library Materials, 41st Meeting, New York, 1996. Disponível em: <http://slis.lsu.edu/faculty/

McCarthy/cdromsalalm.html> Acesso em: 10 de fevereiro de 2005.

MCCARTHY, C. Levantamento geral da automação de bibliotecas no Brasil. *Revista de Biblioteconomia de Brasilia*, v. 18, n. 2, p. 51-57, 1990.

MCCARTHY, C., SCHMIDT, S. Inovação e mudança tecnológica nas bibliotecas brasileiras: a década de noventa. IN: CONGRESSO Brasileiro de Biblioteconomia e Documentação, 17., Belo Horizonte, 1994.

MORENO, Juliana Kiyomura. USP disponibiliza parte de seu acervo de obras raras para pesquisa on-line. Agencia USP de Notícias. 2003 nov. 04. Disponível em: <http://www.usp.br/agen/repgs/2003/pags/241.htm> Acesso em: 11 de fevereiro de 2005.

PASTORE, M. More than 8.6 million online in Brazil". CyberAtlas, 2000. Disponível em: <http://cyberatlas.internet.com/big_picture/geographics/article/0,,5911_463041,00.html> Acesso em: 3 de fevereiro de 2005.

PASTORE, M. Net continues growth in Latin America. CyberAtlas, 2000. Disponível em: <http://cyberatlas.internet.com/big_picture/geographics/article/0,,5911_323391,00.html> Acesso em: 3 de fevereiro de 2005.

Programa Biblioteca Nacional Sem Fronteiras. <URL: http://www.bn.br/fbn/bibsemfronteiras/]> Acesso em 21 de janeiro de 2005.

Projeto: a BibVirt - Biblioteca Virtual do Estudante Brasileiro". 2001. Disponível em: <http://www.bibvirt.futuro.usp.br/abibvirt/projeto.html> Acesso em: 10 de fevereiro de 2005.

PROSSIGA em números. <URL: http://www5.prossiga.br/estatistica/tabelas/> Acesso em: 18 de janeiro de 2005.

SANTANA, P. H. A., PACKER, A. L., BARRETTO, M. Y. Servidor de enlaces: motivação e metodologia. *Ciência da Informação*, v. 30, n. 3, p. 48-55, 2001. Disponível em: <http://www.ibict.br/cionline/300301/3030701.htm> Acesso em: 3 de fevereiro de 2005.

SONEGO, Dubes. A internet caminha para a maturidade. *Meio & Mensagem*, 13 de setembro de 2004.

Statistics of www.scielo.br. <URL: http://www.scielo.br/site_usage/scieloBR-2003.htm> Acesso: 18 de janeiro de 2005.

UNICAMP. Retrato atual da Unicamp em números. 2003. Disponível em: <http://www.unicamp.br/unicamp/a_unicamp/a_unicamp_atuacao_retratoatual.html> Acesso em: 11 de fevereiro de 2005.

Universidade de São Paulo. "A USP hoje". Disponível em: <http://www2.usp.br/portugues/ausp/sobreausp/sobreausp.htm> Acesso em: 18 de janeiro de 2005.

WLADIMIR, J.A., FERNÁNDEZ-JURICIC, E. Regional network raises profile of local journals. *Nature*, v. 415, p. 471-472, 2002.

Web semântica e a gestão de conteúdos informacionais

*Maria Luiza M. Campos**
*Maria Luiza de Almeida Campos/UFF***
*Linair Maria Campos/NCE-UFRJ****

1. Introdução

A Web é hoje considerada o maior repositório de informações dos mais variados domínios de conhecimento, tendo apresentado um desenvolvimento vertiginoso desde sua criação. Suas características de liberdade de publicação, autonomia das fontes e controle descentralizado fizeram com que uma grande diversidade de recursos fosse crescentemente disponibilizada, mudando significativamente o comportamento de seus usuários e ampliando o perfil de sua utilização. O que pode hoje ser observado é uma significativa dependência dos serviços prestados via este novo ambiente de interação, muito além do objetivo inicial de publicação de documentos interligados via uma rede hipertextual de associações. Além de navegar através de ligações entre os documentos, os usuários da Web podem ainda fazer buscas por produtos, verificar situação de sua

*Departamento de ciência da computação/IM. Universidade Federal do Rio de Janeiro
mluiza@nce.ufrj.br
**Departamento de ciência da informação Universidade Federal Fluminense.
mlcampos@nitnet.com.br
***Núcleo de computação eletrônica. Universidade Federal do Rio de Janeiro
linairmc@.nce.ufrj.br

conta bancária, consultar horário de vôos, comprar bilhetes de viagem, para citar apenas alguns dos serviços disponíveis.

A proposta da Web semântica[1] pretende fazer com que essas atividades possam ser executadas não só por usuários, mas também por programas (ou *pela máquina*, como preferem alguns autores). Esses programas ajudariam a automatizar tarefas mais complexas e a decidir, ao longo de atividades intermediárias, qual serviço utilizar, além de obter e fornecer informações, quando necessário, até completar a tarefa especificada. Para que programas possam se comunicar com programas através de uma sofisticada cadeia de interações, eles precisam ser capazes de interpretar tais informações e de julgar a adequação dos serviços a utilizar, eventualmente envolvendo o usuário, quando dele depender qualquer decisão neste processo.

A Web semântica encontra-se nos seus primórdios, e, por isso mesmo, é ainda encarada com ceticismo por muitos. No entanto, cada vez mais surgem abordagens e tecnologias buscando melhorar sua capacidade de prover suporte à identificação, localização, entendimento e uso de seus recursos. Baseada no amplo uso de padrões e arquiteturas extensíveis, capazes de se adaptarem à dinamicidade dos requisitos deste meio, a Web semântica vem trazendo novas perspectivas ao compartilhamento, intercâmbio e integração de recursos, aumentando seu grau de automatização e melhorando significativamente o atendimento às necessidades de seus usuários.

Implícitos a todas essas questões relativas ao tratamento e acesso a recursos na Web, procedimentos relacionados à gestão de conteúdos informacionais se apresentam. Esses procedimentos envolvem operações de classificação, indexação e compatibilização de informações para permitir que recursos mais diversos possam ser agregados, para posterior busca, recuperação e uso efetivos.

A gestão de conteúdos (Heflin; Hendler, 2000) é um conceito recente, que surge no âmbito da ciência da computação, para dar conta do gerenciamento de informações de sistemas corporativos, possibilitando sua organização e acesso. Entretanto, este conceito está diretamente relacionado às atividades de tratamento e recuperação de informação, velhas conhecidas no ambiente da ciência da informação.

Nesta perspectiva, este artigo apresenta as idéias básicas e mecanismos utilizados pela Web semântica, discutindo seus papéis e inter-rela-

ções. Analisa ainda a relação entre estas iniciativas e os procedimentos de organização e representação de domínios, uniformização e padronização, oriundos de áreas como a ciência da informação, com vistas à gestão de conteúdos. É dentro deste contexto que se discutem as mudanças e tendências nesta área, em especial nas novas abordagens aplicadas a bibliotecas digitais.

2. Uma nova geração da Web: a Web semântica

A Web atual foi criada tendo como foco a ampla disponibilidade de conteúdos publicados para consumo por seus usuários. A evolução da Web levou a que cada vez mais estes conteúdos fossem explorados também por aplicações, em um ambiente intensivo de troca de informações, que, no entanto, até recentemente, tem dependido de acordos e predefinições por parte de seus usuários, de forma a garantir a interoperabilidade entre essas aplicações.

A proposta de uma Web semântica busca minorar ou eliminar essa dependência, aumentando a expressividade dos descritores associados aos recursos disponíveis, permitindo automatizar a colaboração entre aplicações e entre estas e seus usuários, através de buscas mais efetivas, integração e interoperabilidade entre aplicações. Se na primeira geração da Web, os documentos eram apenas ligados entre si, deixando para o usuário o papel de interpretar a natureza destas ligações, a nova geração da Web pretende ter ligações semanticamente mais ricas, capazes de dar suporte às aplicações em suas negociações.

Segundo Berners-Lee, Hendler e Lassila (2001):

> A Web semântica representa a evolução da Web atual. Ela visa fornecer estruturas e dar significado semântico ao conteúdo das páginas Web, criando um ambiente onde agentes de *software* e usuários possam trabalhar de forma cooperativa.

Neste novo contexto, a Web será capaz de representar associações entre recursos que, em princípio, poderiam não estar relacionados. Para isso, computadores necessitam ter acesso a conhecimento *consensual* de domínios, através de descritores (metadados) – criados segundo pa-

drões estabelecidos de informações. Além disso, devem também dispor de conhecimento *específico* de cada domínio representado através de redes de conceitos e regras definidas sobre eles. A exemplo de seus precursores na área de inteligência artificial, os idealizadores da Web semântica reconhecem a importância da representação do conhecimento de um domínio, colocando as abordagens e mecanismos daquela área em contato com os novos requisitos e infra-estrutura da Web. Surge então, renovado, o conceito de ontologias (Guarino; Giaretta, 1995), apresentadas como tecnologia chave para a geração da Web semântica, correspondendo a um entendimento comum e compartilhado sobre um domínio capaz de ser comunicado entre homens e aplicações. Nesse sentido, várias iniciativas, como as desenvolvidas pelo W3C, buscam por meio da criação de padrões, arquiteturas de metadados, ontologias e serviços de inferência, dentre outros, a melhor forma de tornar a informação também compreensível pela máquina (Passin, 2004).

3. A descrição e representação de conteúdos informacionais a serviço da gestão de conteúdos na Web

No domínio teórico/metodológico da Web semântica, a gestão de conteúdos se insere como um conceito em construção, contemplando o gerenciamento das informações de um sistema/rede através do tratamento de conteúdos semânticos. Para que esse tratamento ocorra com qualidade, são de fundamental importância investigações relativas a padrões de intercâmbio, controle de linguagem e modelos de representação, através de metadados, vocabulários, ontologias, tesauros, taxonomias, entre outras.

A gestão de conteúdos envolve uma série de procedimentos de uniformização e padronização de informações, baseado numa visão integradora. Nesse contexto, dois conceitos são considerados essenciais: o de organização e o de comunicação.

O conceito de organização pressupõe procedimentos classificatórios. Tais procedimentos possibilitam o agrupamento e a recuperação de informações de acervos diversos. Essas informações podem estar em forma estruturada (bancos de dados, por exemplo) e não estruturada (textos

integrais, por exemplo). Esses procedimentos classificatórios, além de evidenciar os contornos de atuação de uma instituição/organização, facilitam os processos de seleção e de tratamento de informações. O produto desse processo classificatório se apresenta como um mapa de conteúdos – taxonomias – das atividades das organizações que as produzem. Nesse contexto, as taxonomias podem ser consideradas como ferramentas que possuem a função de organização lógica de conteúdos informacionais.

Uma taxonomia é uma lista de categorias de assunto estruturada. Não inclui a definição de tópicos, mas somente apresenta as relações hierárquicas entre os tópicos, como uma parte sistemática de tesauros. Atualmente, o desenvolvimento de taxonomias tem sido um dos pilares da gestão da informação e do conhecimento.

O uso de taxonomias permite que se estabeleçam padrões de alto nível para a ordenação e classificação de informações através de mecanismos de herança, além de permitir que as organizações possam reconhecer e relacionar atividades agregadoras de valor, diminuindo esforços na produção e utilização do conhecimento. O conceito de herança é um dos conceitos mais poderosos no desenvolvimento de *software*. As máquinas podem compreender corretamente relacionamentos de generalização e especialização entre as entidades atribuindo propriedades às classes gerais e então assumindo que as subclasses herdam estas propriedades.

Entretanto, para que uma taxonomia possa verdadeiramente funcionar como um mecanismo que permite a integração das informações, é necessário que exista interoperabilidade entre os diversos sistemas de uma instituição. Isto significa promover a capacidade dos sistemas potencializarem oportunidades de intercâmbio e reutilização de informações, interna ou externamente.

Nesse contexto, a comunicação deve ser entendida como uma série de procedimentos que permite a transmissão de conteúdos informacionais, a partir de uma visão integrada desses conteúdos. Insere-se, nesse domínio, a importância de ações como definição de metadados e construção de terminologias padronizadas. Essas ações viabilizam o tratamento e a recuperação das informações no contexto da Web semântica.

Metadados (Moura e outros, 1999), no ambiente da Web, são em geral entendidos como elementos de descrição/definição de dados armaze-

nados em sistema computadorizado. Organizados em padrões propostos para comunidades específicas, permitem que se determine o conteúdo de cada campo, através de procedimentos como padronização do nome do campo, descrição do conteúdo do campo, tipo de dados, formato etc. A finalidade principal dos metadados é documentar, através de elementos descritores, qualquer tipo de recurso disponível na Web, com o objetivo de permitir comunicabilidade e interoperabilidade entre sistemas. A adoção de padrões de metadados permite com mais facilidade o estabelecimento de mecanismos de importação e exportação de informações, assim como a criação de uma visão integrada dos dados de uma organização. Além disso, permite que agentes inteligentes não somente possam intercambiar informações, mas que possam transferir para um sistema um conhecimento semântico estruturado que é à base da Web semântica.

Por outro lado, é necessário também um meio que viabilize a comunicação não somente entre os sistemas, mas entre o sistema e o usuário desse sistema. Este meio é a linguagem. A linguagem, definida como língua em ação, é composta por pelo menos três objetos, ou seja: a etiqueta lingüística, o conteúdo conceitual e um contexto comunicacional. No espaço da Web semântica esses três elementos são de importância capital para que uma comunicação possa ser estabelecida, permitindo assim acessibilidade às informações disponibilizadas.

As etiquetas lingüísticas que deverão fazer parte do mapa de conteúdo de um *site*, assim como aquelas utilizadas nos mecanismos de busca, são determinadas por uma atividade terminológica. Para cada etiqueta lingüística deve existir um único conteúdo conceitual, dentro de um dado contexto, para possibilitar comunicação, evitar ruído e garantir precisão e qualidade na recuperação das informações. Nesse sentido, uma série de princípios teóricos e metodológicos é adotada para garantir a harmonização de conceitos e termos. (Wuester, 1981; Dalhberg, 1978; Gomes; Campos, 1996)

O produto desta atividade é um vocabulário controlado. No âmbito da Web semântica este procedimento de controle de vocabulário vem sendo representado por diversos instrumentos como tesauros e ontologias.

Tesauro é um vocabulário controlado e dinâmico de termos relacionados semântica e genericamente, cobrindo um domínio específico do conhecimento, funcionando como um dispositivo de controle terminológico

usado na tradução da linguagem natural dos documentos, dos indexadores ou dos usuários para uma linguagem do sistema (linguagem de documentação, linguagem de informação) mais restrita (Unesco, 1973, p. 6). Os tesauros são estruturas sistemáticas de conceitos. Essas estruturas conceituais apresentam termos e relacionamentos entre termos. Os relacionamentos expressos no tesauro podem ser classificados em hierárquicos (gênero/espécie e partitivos), associativos e de equivalência, representados respectivamente pelas siglas TG/TE; TGP/TEP; TA/TA; USE/UP (Campos, 2001).

A partir da década de 90, com o desenvolvimento de estudos que relacionam tesauros a estruturas terminológicas, começam a surgir tesauros que agregam, além de uma parte sistemática e uma parte alfabética, definições para cada termo. Estes tesauros são denominados de tesauros conceituais e seguem a vertente metodológica européia, tendo como princípio a Teoria da Classificação Facetada e a Teoria da Terminologia (Campos; Gomes, 2004).

Diferentemente dos tesauros, que ao surgirem, em finais da década de 60, tinham a finalidade de servir de instrumento para a padronização da linguagem de um serviço de informação, as ontologias ganham expressão no âmbito da Web semântica, já na década de 90, visando a permitir interoperabilidade semântica entre sistemas. Entretanto, ontologias e tesauros possuem pontos convergentes no que tange à função de serem instrumentos que visam padronização terminológica e à recuperação de informação.

Muita confusão se tem feito em torno do conceito de ontologia, que não pode ser considerado somente como um vocabulário controlado. Uma ontologia possui informações de natureza distinta, ou seja, terminológica – possui um conjunto básico de conceitos e relações – e assertivas aplicadas aos conceitos e relações que constituem um conjunto de axiomas (diferentemente de um tesauro).

Embora a construção de ontologias não seja uma tarefa trivial, em muitos casos ela pode partir de vocabulários já existentes, como os tesauros. Estes, se bem estruturados, podem fornecer o ponto de partida para a organização dos conceitos em hierarquias de classes e subclasses, bem como a denominação dos termos associados aos conceitos representados. Para o melhor aproveitamento do tesauro, espera-se que sua

concepção parta da definição de conceitos, como aqueles denominados tesauros conceituais, e não da simples hierarquização de termos usados na linguagem natural. Privilegiam-se, desta forma, os aspectos do significado, uma vez que o foco no conceito permite evidenciar melhor as relações entre os termos representados na ontologia.

A construção de ontologias, no entanto, envolve mais aspectos do que a construção de relações entre termos, como por exemplo, a identificação das diferentes relações entre os conceitos representados e a definição de condições necessárias e suficientes para identificá-los. Estas últimas contribuem para inferências mais ricas do que as fornecidas pelas relações entre conceitos. Por exemplo, se uma laranjeira é uma árvore frutífera, podemos inferir que laranjeira é também uma árvore, assumindo-se que a hierarquia de classes seja organizada como: árvore / árvore frutífera / laranjeira. Por outro lado, vamos supor que definimos que uma *árvore tropical* tem como condições necessárias e suficientes os atributos de ser árvore e ser originária dos trópicos, e que a laranjeira possui esses atributos. Nesse caso, podemos então inferir que laranjeira é também uma árvore tropical, independentemente da hierarquia onde esteja situada na ontologia.

Uma ontologia tem por objetivo prover uma base semântica para esquemas de metadados, facilitando a comunicação entre sistemas e agentes. Desta forma, ontologias estabelecem fundamentos de significados conceituais sem os quais a Web semântica não seria possível, devido à heterogeneidade dos conceitos representados (Jacob, 2003).

A heterogeneidade tem sido identificada como um dos problemas mais importantes e difíceis de serem tratados. Ela envolve a interoperabilidade e cooperação entre múltiplas fontes de informação, retratando diferenças sintáticas, semânticas e estruturais entre sistemas.

A heterogeneidade semântica representa atualmente o maior empecilho para interoperabilidade, representando um grande desafio para integração de informações na Web. Para tratar esse problema é preciso buscar uma linguagem capaz de representar conhecimento e regras, além de inferir novos dados.

Isso se dará a partir de inter-relacionamentos entre ontologias específicas de domínios, que têm como premissa: o uso racional de metadados, para descrição de dados de forma homogênea; o uso sistemático de ontologias, preenchendo a lacuna entre fontes de dados heterogêneas; e

a utilização de associações semânticas, tratando a interoperabilidade entre domínios (Adams, 2002).

4. A Web semântica e o potencial informativo de seus padrões

A Web semântica pretende criar uma nova abordagem para o compartilhamento de recursos, buscando melhorar extraordinariamente os serviços e produtos nela baseados. Ela cria um cenário onde os agentes de *software* desempenham tarefas diárias para os usuários finais, fazendo uso de classificações/taxionomias, metadados, vocabulários estruturados e ontologias.

A Web semântica pretende tornar mais fácil a vida das pessoas respondendo a várias perguntas. Por exemplo, um usuário poderia digitar 'Qual o melhor programa de pós-graduação sobre Gestão de Informação na área da Região Sudeste no Brasil?' Um agente inteligente correria pela Web, compararia a pontuação das universidades de acordo com as avaliações da Capes ou outras e traria uma lista de nomes. O agente inteligente então apanharia o formulário de inscrição e os dados sobre auxílio financeiro e informação do(s) melhor(es) programa(s) de pós-graduação.

Para viabilizar essas facilidades é necessário o uso de padrões e metodologias voltados para fornecer maior clareza para o significado da informação manipulada, acarretando o acréscimo de uma camada extra de infra-estrutura, que faz uso de metadados mais expressivos, à atual HTML.

Na Web atual são utilizadas *tags* META do HTML para inserir informações no cabeçalho (*header*) de uma página Web, associando dados padronizados simples tais como (Figura 1): o autor da página, uma descrição textual do documento e uma lista de palavras-chave. Estas últimas podem ser utilizadas pelas máquinas de busca para indexar a página, juntamente com palavras extraídas do seu título e corpo.

```
<head>
<title>Macaxeira</title>
<meta NAME="description" content=" Comidas Típicas do Nordeste ">
<meta NAME="keywords" content="comida típica, nordeste, macaxeira">
<meta NAME="author" content="Maria Luiza">
</head>
```

Figura 1 – Exemplo do uso de tags META no cabeçalho de uma página HTML

Os descritores permitidos pelas *tags* META são um tipo limitado de metadado, uma vez que fornecem pouca expressividade para descrição do assunto tratado no documento ao qual estão associadas. Para melhorar essas limitações, iniciativas mais abrangentes de padronização foram adotadas, como é o caso do padrão Dublin Core (Dublincore, 2004), desenvolvido nos EUA, com o objetivo de propor um conjunto de elementos padronizado, voltado para a descrição de recursos, incluindo páginas HTML, na Web.

O padrão DC é formado por 15 elementos *(title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights)*, que podem ser acomodados dentro das *tags* META do HTML (Figura 2).

```
<head>
<title>Macaxeira</title>
<meta NAME="DC.description" content=" Comidas Típicas do Nordeste ">
<meta NAME="DC.subject" content="macaxeira">
<meta NAME="DC.language" content="pt">
</head>
```

Figura 2 – Exemplo do uso de *tags* META no cabeçalho de uma página HTML

O DC, embora forneça descritores extensíveis, ainda não permite descrever de forma expressiva os diferentes recursos e conteúdos disponíveis na Web, tendo em vista os agentes de *software*. Por exemplo, não é possível associar ao autor de um livro outros recursos na Web, tais como sua página pessoal, seu *e-mail* e a página da instituição a que ele pertence.

Entretanto, mesmo que padrões como o DC não tenham resolvido as questões de expressividade objetivadas pela Web semântica, um aspecto que não deve ser ignorado é que os metadados padronizados e os vocabulários estruturados tornam mais fácil às bases de dados comunicarem-se entre si, levando em conta o grande problema da Internet atualmente, que é a fragmentação dos dados e a natureza espontânea da descrição de conteúdos. O uso de uma terminologia padronizada possibilita um tratamento automático mais refinado das informações descritas, garantindo consistência e harmonização entre o termo empregado e seu conteúdo informacional. Dessa forma, esse tem sido o caminho seguido pelas iniciativas e padrões ligados à Web semântica, embora muitos as-

pectos devam ser levados em consideração ao definir tais vocabulários padrão, como veremos a seguir.

Com a Web semântica, agentes podem descobrir o significado de uma página Web seguindo ligações hipertextuais dos documentos Web a vocabulários controlados específicos. Por exemplo, as ontologias oferecem referências cruzadas, de modo que um agente compreenda que 'macaxeira', 'aipim' e 'mandioca' constituem diferentes expressões do mesmo conceito. Ainda, as ontologias permitem descobrir que dois conceitos são equivalentes caso ambos compartilhem as mesmas características necessárias e suficientes para defini-los. Por exemplo, se uma pizza doce é definida como sendo uma pizza que tem uma cobertura doce, e se uma pizza possui cobertura de chocolate, e ainda chocolate é definido como sendo doce, então a ontologia permite inferir que uma pizza de chocolate é uma pizza doce.

Para atingir os objetivos pretendidos pela Web semântica, espera-se anotar os recursos disponíveis com metadados expressivos, como os fornecidos pelas ontologias. Além disso, para permitir inferências no nível pretendido, é necessário que as ontologias sejam representadas em uma linguagem compatível com padrões em uso na Web, de modo que sua aceitação e extensão sejam facilitadas, mas com um nível de formalismo tal que permita processamento pelos agentes. Complementarmente, é importante que as ontologias sejam planejadas através do uso de algum método que venha a ajudar a definição e a organização dos conceitos, sempre tendo em mente que a sua construção é uma tarefa multidisciplinar, envolvendo não só as técnicas para sua elaboração, como também o conhecimento do seu domínio alvo.

Várias iniciativas têm sido adotadas de forma interdependente para permitir a concretização da Web semântica. Elas partem do uso do XML[2], que fornece uma estrutura sintática padrão para descrever dados e vêm sendo construídas em busca de maior flexibilidade, maior expressividade semântica e interoperabilidade entre recursos e aplicações.

O uso do XML tem se difundido rapidamente na Web, por ser um padrão simples, em formato texto, com uma estrutura bem definida e facilmente extensível. Porém o XML está no nível da interoperabilidade sintática, pois a semântica das suas representações ainda é implícita, o que pode gerar ambigüidade ao descrever conceitos de um domínio. Por

exemplo, o elemento <identificacao> em um esquema XML pode ser usado para representar o mesmo recurso que o elemento <identidade> em outro esquema, sem que seja possível para a máquina inferir automaticamente que eles são equivalentes. Isso se dá porque não existe uma interpretação geral imposta pelo padrão para as *tags* que descrevem os elementos. Essa interpretação deve ser negociada pelos que usam os documentos. Na Web semântica, entretanto, espera-se que os agentes de *software* sejam capazes de interpretar a descrição de um recurso de forma independente, sem ambigüidade, de modo a poder localizá-los, combiná-los com outros, se for conveniente, e tomar decisões baseadas em conhecimentos que podem ser inferidos a partir dos significados interpretados.

Nesse contexto, surge o RDF[3], que faz uso do XML para permitir descrever e pesquisar recursos de forma mais flexível, mediante representação de relacionamentos entre estes, na forma de triplas compostas por objeto-atributo-valor, sendo que um valor pode ser um literal ou outro recurso. Através do RDF, é possível representar afirmações simples, como, por exemplo: "o autor da música *Beatriz* é *Chico Buarque*" e "Chico Buarque possui *e-mail* chico@xxx.com.br". Estas afirmações podem também ser representadas em forma de *grafo*. Neste caso, recursos são representados por elipses, enquanto a associação de valores literais a estes é representada por uma caixa retangular, como mostra a Figura 3.

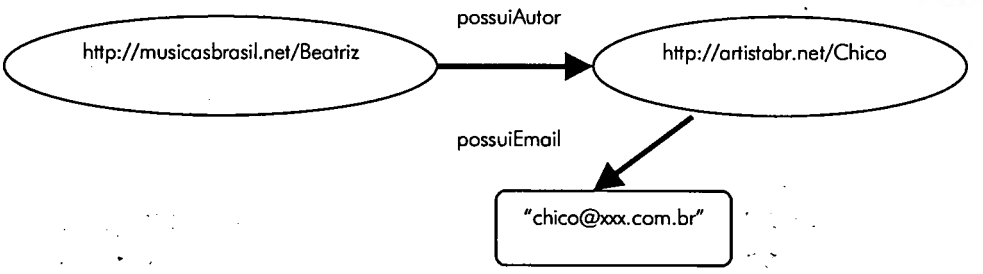


Figura 3 – Grafo RDF representando *Chico Buarque* como autor de *Beatriz*, e seu *e-mail*

Uma outra vantagem do RDF é que ele é facilmente extensível pelo uso de outros esquemas de representação de domínios, que podem ser usados para ampliar a semântica implícita que ainda se encontra em sua proposta. Nesse sentido, o RDF *Schema*[4] (RDFS) vem se juntar ao padrão

RDF, fornecendo a este um nível superior de abstração que permite que conceitos primitivos, ou tipos de objetos sejam definidos, cuja semântica pode ser entendida por agentes de *software* (Hui e outros, 2003). Desta forma, através do RDFS é possível definir uma terminologia para representar conceitos como *recurso*, *classe (tipos de recurso)*, *subclasse* e *propriedade (atributos das classes)*, que podem ser usados para expressar organizações hierárquicas de conceitos e suas relações. Por exemplo, no caso da figura 3, o termo *propriedade* é usado para definir que *possuiAutor* e *possuiEmail*, são relacionamentos, deixando-se claro de que forma *propriedade* deve ser usado e o que ele significa. De maneira análoga, outros esquemas podem ser usados ainda para contextualizar o significado de propriedades como *possuiEmail* abrindo caminho para a descoberta de novos conhecimentos através de inferências sobre afirmações cuja semântica agora passa a ser explícita. Dessa forma, o RDFS vem apoiar não só a utilização de vocabulários diversos para diferentes domínios, como também dar um sentido consensual a alguns elementos básicos que podem ser explorados pelos agentes de *software* que estejam de acordo com esta semântica, para interpretar de forma inequívoca as afirmativas feitas sobre os recursos por eles manipulados.

Entretanto, o RDFS não foi concebido como sendo uma linguagem para construção de ontologias, e, dessa forma, apresenta algumas limitações que foram superadas por linguagens como o DAML+OIL[5] e, mais recentemente OWL[6], que vem sendo adotada como padrão para elaboração de ontologias na Web.

OWL possui uma terminologia mais poderosa para formalizar a definição de conceitos, estendendo o RDFS para descrever características tais como condições necessárias e suficientes (que servem para indicar os atributos mínimos que definem um conceito), disjunção (que serve para indicar que dois ou mais conceitos são coisas distintas), interseção e união, e ainda cardinalidade de relações, relações de equivalência, inversas, simétricas, dentre outras. Com base nesse formalismo mais sofisticado, ferramentas chamadas de *reasoners* são capazes de fazer inferências mais poderosas, bem como validar a ontologia de acordo com as afirmações que são feitas ao se definirem seus conceitos. Por exemplo, podemos definir um vinho branco por meio das condições necessárias e suficientes (representadas pela terminologia de *EquivalentClass*) de ser um

vinho e ter a cor branca (*e* é representado por intersectionOf, no caso unindo as duas condições Vinho e a restrição de sua cor). O que significa que se se a *champagne* é definida como um vinho, e se tem como atributo de cor "branco", então pode-se inferir através dessa ontologia que *champagne* é um vinho branco.

```
EquivalentClass (VinhoBranco
  intersectionOf(
    Vinho
    Restriction (corVinho
  hasValue("branco"))))
```

Figura 4 – Definindo um possível conceito de vinho branco em OWL

Além de descrever e localizar recursos, a Web semântica envolve a realização de tarefas, de forma sincrônica ou não. Nesse cenário, é possível, por exemplo, uma pessoa em férias tirar uma foto com a câmera de seu celular, usar o aparelho para procurar por um agente de impressão de fotos digitais, fazer o pagamento da impressão através de débito em cartão de crédito, cuja validação é feita por um outro agente, e solicitar a entrega das fotos para a sua residência. As tarefas desempenhadas pelos agentes podem ser executadas por serviços Web, transformando assim a Internet em uma fonte de serviços distribuídos interoperáveis, que podem ser executados a qualquer hora, de forma individual ou combinada, e que são apoiados por uma arquitetura padrão (Services Open Architecture-SOA) que prevê mecanismos para a sua descoberta (Universal Description, Discovery and Integration-UDDI), descrição (Web Services Description Language-WSDL) e troca de mensagens (Simple Object Access Protocol-SOAP) de maneira independente de plataforma e de linguagem de programação[7].

Entretanto, os padrões ligados aos serviços Web possuem limitações que dificultam a sua busca automatizada. Nesse contexto, várias iniciativas têm sido conduzidas para melhorar diferentes aspectos da arquitetura padrão de serviços Web, adicionando maior semântica a esta. O objetivo é originar um novo contexto para o uso desses serviços, que favoreça automatizar não só a sua descoberta, como também seu uso de forma combinada. Uma dessas iniciativas é a OWL-S (Martin *et al.*, 2003),

uma ontologia para descrição de serviços Web em termos de seus aspectos funcionais e não funcionais, que pode ser estendida para descrever aspectos específicos de cada área de aplicação dos serviços.

A partir destas iniciativas, apoiando as idéias e padrões associados à Web semântica, desenvolveu-se uma infra-estrutura de *software* capaz de apoiar seu uso e dar suporte à construção de uma nova geração de aplicações. Dentre estas ferramentas, incluem-se sistemas e interfaces de programação (as chamadas *API*s) que ajudam na descrição e anotação de conteúdos (McBride, 2005), na validação e construção de ontologias (Horrocks, 1998), na descoberta em tempo real de novos conhecimentos a partir das definições formais aí existentes (Haarslev e Möller), na busca e recuperação de recursos (incluindo-se aí, serviços Web) (Fensel e Bussler, 2002). Uma das ferramentas mais populares para criação e manipulação de ontologias é o Protégé[8], que dá suporte à representação em OWL assim como outras funcionalidades para permitir a sua visualização e validação. Uma lista detalhada de ferramentas e iniciativas ligadas à Web semântica pode ser encontrada na página Web da semanticweb.org[9].

5. Bibliotecas digitais e a gestão de conteúdos na era da Web semântica

A variedade de formatos e o volume de dados têm crescido de maneira acelerada nos últimos anos. Hoje em dia um documento comumente já é constituído não apenas de texto, mas de gráficos, imagens, sons e vídeos. Esses documentos se encontram nos mais diversos locais: em bancos de dados, em sistemas de arquivos, espalhados pela Web, ou ainda armazenados em bibliotecas digitais. Estas últimas, fornecendo vantagens tais como acesso remoto e distribuído, facilidade e rapidez de busca, e referências cruzadas a outros documentos. Entretanto, estes documentos possuem uma característica comum: devido à sua diversidade de formato, as diferentes plataformas, meios e locais de armazenamento, e também à sua atualização dinâmica, eles necessitam de um novo modelo, mais expressivo e ágil para seu gerenciamento.

Ainda sem um foco na Web semântica, Hodge (2000) apresentou interessantes considerações sobre o papel de sistemas de organização de conhecimento no contexto de bibliotecas digitais, salientando que

estes sistemas vão bem além do objetivo inicial da organização dos acervos, podendo dar suporte à interligação dos recursos, a mecanismos de recuperação mais sofisticados explorando as ricas relações entre os conceitos representados.

Essas características se mostram alinhadas com os objetivos da Web semântica, onde as tecnologias de descrição da informação têm evoluído para prover maior formalismo de representação e conseqüentemente maiores possibilidades de recuperação de informação relevante, por meio do uso de agentes inteligentes de *software* capazes de utilizar sofisticados mecanismos de inferência.

É interessante observar que hoje tipos de conteúdo diversos podem ser descritos. Por exemplo, em ferramentas de busca como o Google[10], pode-se fazer buscas por imagens. Entretanto, o tipo de retorno obtido é, em geral, impreciso e genérico. Não é possível, com a tecnologia tradicional, obter respostas a questões, como: localize imagens de tecidos que tenham diagnóstico de câncer com tamanho de 20 mm.

Nesse cenário, as ontologias têm um importante papel a cumprir: o de fornecer a base semântica para que mecanismos sofisticados de descrição e recuperação de dados possam atuar, levando em conta a heterogeneidade não só dos documentos em si, mas também dos vocabulários ou metadados utilizados para descrevê-los.

Os ambientes de gerenciamento de conteúdos da Web semântica levam em conta esses aspectos e fazem uso de processos que intermedeiam essas diferenças de forma automatizada, utilizando recursos tais como busca mediada por ontologias diversas e serviços Web. Como exemplo de tais ambientes, podemos citar o JeromeDL (Kruk e Synak, 2004) e o MarcOnt[11], este último ainda em implementação.

JeromeDL é uma biblioteca digital de código aberto que permite a descrição de recursos em RDF, sendo compatível com o padrão Dublin Core. Sua principal vantagem está no uso de buscas semânticas baseadas em ontologias, em vez de palavras-chaves, melhorando consideravelmente a precisão da busca. Além disso, permite a descrição da estrutura do documento, por exemplo: capítulos, intervalos de páginas, partes, anexos. Uma outra vantagem é a utilização de SOAP como protocolo de troca de mensagens, permitindo integração de recursos de diferentes bibliotecas, com diferentes formatos de descrição.

O projeto MarcOnt é mais ambicioso que o JeromeDL, na medida em que propõe uma arquitetura mais formal e robusta para a descrição de recursos. Um de seus objetivos principais é a integração de bibliotecas digitais, utilizando para isso serviços Web, ontologias em OWL e ferramentas para tradução entre diversos padrões de descrição. OWL é usado com o intuito de permitir inferências e para expressar padrões de descrição tais como o Dublin Core e o MARC21[12], este último um padrão originário da Biblioteca do Congresso americana[13], e em uso por mais de 30 anos em bibliotecas. Além disso, estão previstos recursos de inteligência artificial tais como mineração de dados e algoritmos evolucionários (capazes de aprendizado através de modelos). O projeto teve início em meados de 2004, esperando-se que os primeiros resultados sejam divulgados em um ano.

Como podemos observar, a Web semântica está ainda em seus estágios iniciais, mas há evidências de que sua consolidação depende apenas de tempo e do amadurecimento das tecnologias utilizadas. Uma dessas evidências é o investimento crescente de grandes fabricantes, organizações internacionais de padronização e comunidades, com destaque para a área de bibliotecas digitais, na qual problemas de tratamento de informação são cada vez mais críticos, devido ao volume e à complexidade cada vez maior de manipular e integrar informações heterogêneas e distribuídas de maneira dinâmica.

7. Conclusão e tendências

No ambiente da Web semântica, a gestão de conteúdos informacionais amplia suas fronteiras, revendo métodos e abordagens tradicionalmente utilizados, expandindo suas funcionalidades de forma a explorar o potencial das recentes inovações. Em especial, as bibliotecas digitais, representando ilhas de informação de reconhecida qualidade, ganham novo impulso, não apenas com as melhorias nos mecanismos de indexação, recuperação e navegação, mas principalmente na integração e interligação de seus acervos, ampliados diante da possibilidade de conexões automáticas a novos recursos.

A nova Web depende do desenvolvimento de ferramentas que facilitem a geração e a gerência de todo este rico conjunto de descritores, garantindo criação de camadas semânticas por sobre os recursos dispo-

níveis. Os profissionais da informação têm muitas habilidades que os cientistas da computação e outros procuram quando tentam vislumbrar a Web semântica, podendo contribuir decisivamente para sua próxima etapa de desenvolvimento. Como apontado por Alvarenga (2001), a resposta para muitas das questões na área de bibliotecas digitais pode estar na combinação e adaptação de métodos e práticas da biblioteconomia às tecnologias de *software* desenvolvidas recentemente. Não é à-toa que atualmente verificamos uma série de questões comuns que se configuram no fazer destes profissionais, a exemplo de organização de modelos de representação do conhecimento, classificação de recursos informacionais, elaboração de vocabulários e ontologias, organização de informação. Por isso, espera-se maior intercâmbio entre estas comunidades, desde a efetivação de uma terminologia comum até o desenvolvimento de trabalhos conjuntos, especialmente no domínio da gestão de conteúdos na Web semântica.

Notas

[1] Semantic Web Home Page: www.w3.org/2001/sw/

[2] Extensible Markup Language (XML): http://www.w3.org/XML.

[3] Resource Description Framework (RDF): http://www.w3.org/TR/rdf-primer/.

[4] RDF Schema (RDFS): http://www.w3.org/TR/rdf-schema/

[5] DAML+OIL Reference Description: http://www.w3.org/TR/daml+oil-reference

[6] Web Ontology Language (OWL) Reference: http://www.w3.org/TR/2002/WD-owl-ref- 20021112/

[7] Web Services Activity Homepage: http://www.w3.org/2002/ws/

[8] Protege Ontology Editor and Knowledge Acquisition System: http://protege.stanford.edu/

[9] Semantic Web Homepage: http://www.semanticweb.org.

[10] Ferramenta de Busca Google: http://www.google.com

[11] Projeto MarcOnt: http://www.marcont.org

[12] Machine-Readable Cataloguing Standards (MARC): http://www.loc.gov/marc/

[13] American Library of Congress: http://www.loc.gov/

Referências

ADAMS, K. *The Semantic web: differentiating between taxonomies and ontologies*. Online, v. 26, n. 4, p. 20-23, July/August, 2002.

ALVARENGA, L. *A teoria do conceito revisitada em conexões com ontologias e metadados no contexto das bibliotecas tradicionais e digitais*. Datagramazero, v.2, n.6, dez. 2001.

BERNERS-LEE, T. ; HENDLER, J.; LASSILA, O. *The semantic Web*. Scientific American, 2001.

CAMPOS, M. L. A. *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: EdUFF, 2001. 133p.

CAMPOS, M. L. A.; GOMES, H. E.; MOTTA, D. F. *Tutorial de tesauro*, 2004. Disponível em : http://www.conexaorio.com/biti/tesauro.

DAHLBERG, I. A. *Referent - oriented analytical concept theory of interconcept*. International Classification. v.5, n.3, p.142-150, 1978.

DUBLINCORE Initiative, *Dublin core metadata element set, version 1.1*: Reference Description. p. 17-20, 2004. Disponível em: http://dublincore.org/documents/dces/.

FENSEL, D.; BUSSLER, C. *The Web service modeling framework-WSMF*. Electronic Commerce Research and Applications, v.1, n.2, 2002.

GOMES, Hagar E. ; CAMPOS, M. L. A. *Systematic aspects of terminology*. Meta, v.41, n. 2, p. 249-254, jun. 1996.

GUARINO, N.; GIARETTA, P. *Ontologies and knowledge bases: towards a terminological clarification*. Towards very large knowledge bases: Knowledge building and knowledge sharing, pp. 25-32. Amsterdam: IOS Press, 1995.

HAARSLEV, V.; MÖLLER, R. *High performance reasoning with very large knowledge bases: a practical case study*. 17th International Joint Conference on Artificial Intelligence, August 4-10, 2001, Seattle, Washington, USA.

HEFLIN, Jeff & HENDLER, James. *Semantic interoperability on the Web*. Extreme Markup Languages 2000, 2000.

HODGE, G., *Systems of knowledge organization for digital libraries: beyond traditional authority files*. The Digital Library Federation Council on Library and Information Resources, Washington, 2000.

HORROCKS, I. *Using an expressive description logic: FaCT or fiction*? Principles of Knowledge Representation and Reasoning: Proceedings of the Sixth International Conference, p. 636-647, June 1998.

HUI K., CHALMERS, S., GRAY, P., PREECE, A. *Experience in using RDF in agent-mediated knowledge architectures*. Workshop on Agent-Mediated Knowledge Management, Spring Symposium, p. 177-192, 2003.

JACOB, E.K. *Ontologies and the semantic Web*, Bulletin of the american society for information science and technology, April/May, 2003.

McBRIDE, B. *An introduction to RDF and the jena RDF API*, 2005. Disponível em: http://jena.sourceforge.net/tutorial/RDF_API/index.html.

KRUK, S. R; SYNAK, M. *JeromeDL – e-library with semantics*.Technical report, DERI.NUIG - Ireland; Gdansk University of Technology - Poland, 2004. Disponível em: http://www.jeromedl.org.

MARTIN, D.; BURSTEIN, M.; DENKER, G.; J. HOBBS, J; KAGAL, L; LASSILA, O; MCDERMOTT, D; MCILRAITH, S; PAOLUCCI, M; PARSIA, B; PAYNE, T; SABOU, M; SIRIN, E; SOLANKI, M; SRINIVASAN, N; SYCARA, K. *OWL-S 1.0 draft release*. Disponível em: http://www.daml.org/services/owl-s/1.0/, Dec 2003.

MOURA, A. M. C.; CAMPOS, M. L.; BARRETO, C. M. *A survey on metadata for describing and retrieving Internet resources*. WWW Journal, v.1, p. 221-240, fev. 1999.

PASSIN, T.B., *Explorer's guide to the semantic Web*. New York: Manning, 2004, 281p.

UNESCO. Guidelines for the establishment and development of monolingual thesauri. Paris, 1973, 37p.

WUESTER, E. *L'Étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les Sciences des Choses. Textes choisis de terminologie*. I. Fondéments théoriques de la terminologie, Québec, p. 57-114, 1981.

2. Dimensão Tecnológica

Infra-estrutura tecnológica de uma biblioteca digital: elementos básicos

*Silvana Aparecida Borsetti Gregorio Vidotti**
*Ricardo Gonçalves Sant'Ana***

1. Introdução

Para se pensar em uma infra-estrutura tecnológica para uma biblioteca digital, é necessário o entendimento dos conceitos relacionados à expressão "Biblioteca Digital":

– Biblioteca: entendida de forma simplificada como instituição que tem por princípios básicos a seleção, o armazenamento, a representação descritiva e temática, a organização, o acesso, a recuperação e a disseminação de itens documentários aos usuários potenciais.
– Digital: informação em formato digital (em computação – dígitos 0 ou 1)

Utilizando os conceitos de biblioteca e digital, pode-se afirmar que o desenvolvimento de uma biblioteca digital se baseia no planejamento de

* Doutora em educação – UNESP. Professora de tecnologias em informação do Departamento de Ciência da Informação, atua nos cursos de graduação em arquivologia e biblioteconomia e no programa de pós-graduação em ciência da informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista – UNESP. Membro do Grupo de Pesquisa – Novas Tecnologias em Informação. vidotti@marilia.unesp.br
** Mestre e doutorando em ciência da informação – UNESP. Professor de tecnologias de informática do Departamento de Ciência da Informação, atua nos cursos de graduação em arquivologia e biblioteconomia da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista – UNESP. Membro do Grupo de Pesquisa – Novas Tecnologias em Informação. ricardosantana@marilia.unesp.br

uma biblioteca tradicional/convencional, desde o processo de aquisição (compra, digitalização, acesso a outros *sites* e auto-arquivamento), o processamento técnico (catalogação, classificação, indexação – metadados e iniciativa de arquivos abertos), a recuperação (ferramentas de busca), a disseminação (boletins eletrônicos), o atendimento ao usuário (setor de referência digital – meios de comunicação digital e sistemas agentes), até a preservação (itens documentários e dos suportes informacionais). Neste sentido, torna-se necessário um estudo sobre as funcionalidades, as características e os serviços a serem oferecidos, bem como uma política de desenvolvimento de coleções baseada em tipos documentais, conteúdos informacionais e público-alvo, e uma política de preservação.

O planejamento da estrutura digital passa pelas etapas de definição da Arquitetura da Informação de *website*, que culmina com o dimensionamento do sistema de computação (*hardware* e *software*), visando a integração dos serviços e dos conteúdos informacionais para uma interação eficiente do usuário com o ambiente informacional.

Definidos os requisitos básicos de funcionamento de uma biblioteca digital, passa-se ao processo de aquisição, no qual são definidos os conteúdos/recursos informacionais que comporão o acervo digital por meio de compra, assinatura, digitalização, auto-arquivamento e seleção de obras pertencentes a outras bibliotecas digitais, repositórios institucionais, periódicos científicos e/ou outros *websites*. O acervo pode ser composto por recursos multimídia (informações textuais, sonoras e imagéticas) interligados através de informações referenciais e/ou contextuais.

Após a aquisição dos itens documentários digitais, inicia-se o processamento técnico dessas obras com a elaboração de metadados, que são "um conjunto de dados referenciais metodologicamente estruturado e codificado, conforme padrões internacionais, para localizar, identificar e recuperar pontos informacionais de documentos". (Siqueira; Santos, 2004, p. 96). Como exemplo, tem-se: MARC em XML, Dublin Core (DC), Internet Anonymous FTP Archive (IAFA), Summary Objects Interchange Format (SOIF), Text Encoding Iniciative (TEI) e Encoded Archival Description (EAD). Vale destacar que o OAI (Open Archives Initiative) provê uma forma padrão para tornar disponível o acervo digital via Internet, bem como para a coleta de informações de outros acervos

via consultas através do protocolo OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).

O processo de recuperação das informações pode ser baseado em estruturas de diretório que classificam as obras/recursos digitais segundo sistemas de classificação hierárquico predefinidos, como CDD (Classificação Decimal de Dewey) ou CDU (Classificação Decimal Universal), e em estruturas de busca direta via ferramenta de busca, que consiste em percorrer toda a base de metadados à procura das informações que satisfaçam a expressão de busca. Além disso, a recuperação pode ser feita em diferentes bibliotecas digitais e/ou outros *websites* de forma similar a um metapesquisador.

A atividade de disseminação pode consistir na elaboração e no envio de boletins eletrônicos, *e-mails*, listas de discussões que divulguem a incorporação de novos conteúdos, bem como de conteúdos específicos a cada usuário potencial, segundo critérios estabelecidos no momento do cadastro do usuário e em atividades dinamicamente estabelecidas conforme interação do usuário com o sistema.

O atendimento digital ao usuário, conforme serviço de referência, pode ser feito por várias formas de comunicação que vão desde *e-mails* e IRQ, ICQ, *chat – softwares* que têm a característica de serem *on-line*, possibilitando "conversas" virtuais em tempo real –, até sistemas agentes de comunicação.

A política de preservação dos recursos/objetos deve ser enfocada em termos de integridade lógica e física do ambiente informacional, que consiste na instalação de *softwares* e *hardwares* de segurança quanto à questão de acesso e uso, bem como da manutenção e da atualização dos suportes informacionais, considerando a evolução destes suportes e sua vida útil.

Conforme Wainwright (1996, p.2, tradução nossa), reafirma-se, então, que

> a biblioteca digital continua sendo uma biblioteca, com os mesmos propósitos, funções e metas da biblioteca tradicional. A parte digital do termo indica meramente que os materiais são armazenados e acessados digitalmente. Uma biblioteca digital é então muito mais que uma coleção digital, particularmente uma coleção, como é ilustrado no número grandioso de *sites* da World Wide Web na Internet que consiste em informação atual relativamente volátil[1]

2. Arquitetura da Informação de *website*

Os estudos apontam a interligação da Arquitetura da Informação com os processos, métodos e ferramentas utilizados pelas bibliotecas tradicionais, como formas de organização e implantação de ambientes informacionais digitais de busca, interação, promoção e usabilidade de informações, porém com um enfoque específico nos serviços e produtos digitais, de modo a criar sistemas de seleção, descrição, armazenamento, representação, indexação, recuperação e disseminação de informações digitais que possibilitem a construção, o armazenamento e a disseminação do conhecimento.

A expressão "Arquitetura da Informação" foi apresentada pelo arquiteto e desenhista gráfico Richard Saul Wurman, na década de 1960, relacionada às diversas formas como a informação sobre ambientes urbanos poderia ser reunida, organizada e apresentada para públicos distintos. Com isso, Wurman viu a arquitetura como uma ciência e arte de criar uma "instrução para espaço organizado" (Wyllys, 2000, p. 1).

Diversos autores apresentam características da Arquitetura da Informação em *websites* de diferentes naturezas, como McGee e Prusak (1994), Donati, Carvalho e Prado (1998), Rosenfeld e Morville (1998), Straioto (2002), Carrett (2003), Sarmento e Souza, Foresti e Vidotti (2004) e Camargo (2004).

A Arquitetura da Informação descrita por Rosenfeld e Morville (1998) apresenta os seguintes elementos, que podem ser utilizados no desenvolvimento de *website*, e que podem ser aplicados em biblioteca digital:

– Sistema de organização: formas de agrupamento dos conteúdos informacionais – responsável pela estruturação dos conteúdos, nele são definidos os critérios de disposição dos itens informacionais, observando os esquemas e/ou estruturas que melhor satisfaçam a necessidade do usuário, sem comprometer a navegabilidade do *website*.

Os esquemas podem ser dos tipos exatos, ambíguos ou híbridos. Os exatos são caracterizados pela criação de seções bem definidas, e se dividem em: alfabético (apresentação da informação em ordem alfabética), cronológico (informações ordenadas cronologicamente) e geográfico (informações por regiões geográficas). Os ambíguos, em tópicos (tipos de

classificação por assuntos ou tópicos específicos), orientados a tarefas (organiza conteúdos e aplicações em conjuntos de funções e processos), específicos a um público (com acesso restrito ou não), e dirigidos a metáforas (utilizam metáforas comuns ao usuário, que lhe permitem entender itens de informação a serem acessados). Já os híbridos contêm mais de um tipo de esquema, tanto dos esquemas exatos quanto dos ambíguos, e são geralmente encontrados em *websites*, uma vez que um único esquema de organização nem sempre é suficiente para representar o conteúdo informacional. Destaca-se que a utilização de esquemas deve sugerir modelos mentais associados à estrutura cognitiva dos usuários.

As estruturas podem ser do tipo hierárquicas, de base relacional ou hipertextuais. As hierárquicas, geralmente, apresentam uma ordenação de classes por ordem de subordinação (do geral para o específico – *top-down*). As de base relacional possibilitam o acesso por informações de campos específicos ou pela combinação deles, inicialmente em termos específicos e culminando em termos mais genéricos (do específico para o geral – *bottom-up*). As hipertextuais são uma forma não linear de organização (rede informacional).

– Sistema de rotulagem: denominação do conteúdo do grupo informacional - forma de representar um conjunto de informações utilizando uma palavra ou um ícone, de modo a facilitar a recuperação da informação e a navegabilidade do *web site*. Em outras palavras, é a representação ou identificação textual ou iconográfica de um conteúdo específico, objetivando facilitar e tornar familiar uma forma de organização de informações. Possibilita ao usuário decidir qual caminho seguir, permitindo que o mesmo possa identificar-se com a linguagem e com a estruturação do *site*, culminando em um tempo de navegação mais otimizado.

– Sistema de navegação: forma de interação do usuário com o ambiente e com o conteúdo informacional disponível, permitindo ao usuário ir de um ponto ao outro pelo caminho desejado, possibilitando um melhor aproveitamento do tempo de uso ou de acesso. Mantém o equilíbrio entre a movimentação e o oferecimento de opções de *links*

para o usuário. As estruturas de navegação baseiam-se geralmente nas formas hierárquicas, globais e locais.
A estrutura de navegação hierárquica permite o acesso às informações dos níveis mais gerais até os mais específicos. A estrutura de navegação global possibilita o acesso a diferentes tipos de informações e serviços (movimentos laterais – largura), e o acesso a itens específicos de informações (movimentos verticais – profundidade). O sistema de navegação local é um sistema utilizado para complementar o sistema global, permitindo uma navegação entre um conjunto particular de páginas do próprio *website*. A estrutura de navegação *ad-hoc* apresenta *links* inseridos no corpo do texto, que podem ser palavras ou frases, fornecendo informações adicionais sobre um assunto, em uma forma mais textual. Dentre os elementos que permitem a estruturação da navegação destacamos: a barra de navegação, os *frames*, o sumário, os índices e o mapa do *site*.
Assim, conforme Nunes (2000, p.1),

> Arquitetura de navegação define-se como a criação de estruturas estabelecidas por vários *links* que formam um organograma com todas as páginas existentes dentro de um mesmo *site*. Os *hiperlinks* proporcionados pela HTML são verdadeiras "pontes de ligação" entre o conteúdo das páginas de um *site*, e por meio deles, é possível direcionar e estabelecer possíveis caminhos ao visitante. As estruturas de navegação não são exclusivas dos *sites* da Internet e a mesma maneira de funcionamento serve também para CD-ROMs e outras publicações em hipermídia. Excetuando diversos *sites* que colocam o mapa do *site* à disposição do visitante, na maioria das vezes a estrutura de navegação permanece oculta e só pode ser deduzida após algumas páginas internas serem vistas.

– Sistema de busca: possibilita ao usuário a formulação das expressões de busca para a recuperação dos documentos que correspondem à informação desejada. Para a recuperação das informações é necessária uma forma de representação descritiva e temática adequada aos conteúdos dos metadados. Deve-se observar, antes da implementação, a forma como os usuários potenciais realizam essas buscas, e seus diferentes tipos de necessidades informacionais. Os tipos de busca apresentados por Rosenfeld e Morville (1998) são: busca por itens

conhecidos, quando o usuário sabe exatamente o que deseja; busca por idéias abstratas, em que o usuário tem uma vaga noção do assunto; buscas exploratórias, quando o usuário pesquisa para aprender mais sobre um determinado assunto; e buscas compreensivas, em que os usuários desejam todas as informações possíveis sobre o assunto a ser pesquisado.

Além dos sistemas de organização, rotulagem, navegação e busca, vale destacar a usabilidade que enfoca a análise do *design* do *web site*, verificando a possibilidade um pleno funcionamento, identificando e solucionando possíveis erros, tornando-os fáceis de usar em prol do usuário. Pode ser aplicada durante a criação ou na remodelação de um *web site*.

Assim, a usabilidade baseia-se no constante estudo dos usuários potenciais, aplicando os chamados testes de usabilidade; "eles são os instrumentos mais indicados para medir a taxa de sucesso da busca de informações e para pesquisar o comportamento dos indivíduos durante a interação com os sistemas informatizados". (Agner, 2004, p.1). Na realização destes testes, são escolhidos tipos de usuários diferentes, com ou sem conhecimento ou familiaridade com o *website*, e o cruzamento dos resultados obtidos por esses "públicos" diferentes dirá se todas as tarefas realizadas (como, por exemplo, a busca de determinada informação no *site*) foram realizadas com sucesso ou não, apontando as preferências dos usuários.

Portanto, a verificação da usabilidade do *website* faz-se necessária uma vez que, segundo Dias (2003, p.28):

> Um sistema interativo é considerado eficaz quando possibilita que os usuários atinjam seus objetivos. A eficácia é a principal motivação que leva um usuário a utilizar um produto ou sistema. Se um sistema é fácil de usar, fácil de aprender e mesmo agradável ao usuário, mas não consegue atender a objetivos específicos de usuários específicos, ele não será usado, mesmo que seja oferecido gratuitamente.

Estruturado nos elementos da Arquitetura da Informação e baseado nos princípios de usabilidade descritos anteriormente, o *website* deve, portanto, revelar seus serviços, seus conteúdos e suas formas de interação de maneira simples, de modo a gerar uma confiabilidade da instituição ao usuário.

3. Arquitetura tecnológica

Para tornar os serviços e conteúdos da biblioteca digital acessíveis via Internet, devem ser consideradas algumas características do funcionamento da Internet. Essa rede permite que milhões de computadores troquem informações entre si, e isso só foi possível devido à utilização de conceitos como o de pacotes de informações e de endereçamento. Toda informação a ser transmitida é fragmentada em pacotes (partes) e então enviada para seu destino, contendo o endereço do remetente e do destinatário.

Quando o usuário se conecta à Internet, recebe um endereço, que pode ser diferente a cada conexão, já que sua única função é permitir que receba as respostas às solicitações geradas.

Porém, quando o objetivo é fornecer um determinado serviço à rede, como, por exemplo, tornar disponível o conteúdo de uma biblioteca virtual, deve-se obter um endereço fixo e assim poder ser encontrado pelos usuários que se pretende atender.

A Internet utiliza um esquema de endereçamento e nomenclatura chamado endereçamento IP (Internet Protocol). Cada endereço IP é composto por quatro números, cada um com valor entre 0 e 255, compondo, assim, um endereço único em toda a rede para cada ponto ou nó desta rede.

O conjunto de elementos de *hardware* e *software* que formam a estrutura do site compõem uma rede local e para que esta rede local possa se comunicar com outras redes é necessário um equipamento capaz de enviar e receber informações de outras redes, o roteador, que terá seu próprio endereço de rede, conhecido como *gateway*, endereço este que deverá estar registrado em todos os demais equipamentos que fazem parte da rede local, pois é através do roteador que os mesmos poderão se comunicar com a Internet.

As informações que trafegam entre os equipamentos da rede local são identificadas por endereços específicos que podem ser classificados em função do cálculo realizado em conjunto com uma identificação da própria rede, conhecido como máscara de rede. Se o endereço de destino da informação não pertencer à rede local, esta informação será direcionada para o roteador, que irá então direcioná-la para as redes externas e, por conseguinte, para a Internet.

Não é tarefa fácil lembrar esta seqüência de números cada vez que se deseja acessar um determinado endereço, e para simplificar esta tarefa é

utilizado outro esquema de simplificação, *Uniform Resource Locator* (URL), através do qual é atribuído um nome para cada endereço, chamado nome de *host* ou nome de domínio.

O endereço IP, do binário ao domínio, é ilustrado da seguinte maneira: 11001000 10010001 10101011 00000011 à 200.145.171.3 à www.marilia.unesp.br. Para visualizar o endereço IP de um determinado endereço, como a Unesp – Campus de Marília, o usuário pode utilizar, por exemplo, a opção Executar do Windows e digitar o comando: tracert www.marilia.unesp.br. Neste momento, será exibido o endereço IP deste endereço de *host* e o caminho percorrido.

Para que seja possível utilizar este esquema de endereços é preciso registrar seu próprio endereço de IP e seu domínio. No Brasil, é possível registrar seu endereço através de um cadastro junto à Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP. Maiores informações podem ser obtidas no endereço: http://registro.br/index.html, onde é possível verificar se o domínio já existe e realizar todo o processo de registro, que consiste no preenchimento de formulários e pagamento da taxa referente ao registro e ao primeiro período.

A World Wide Web, conhecida também como Web, é um dos serviços disponíveis na Internet. A Web oferece uma interface gráfica com o usuário (*Graphical User Interface* – GUI), colorida e de fácil utilização. O utilitário necessário para acessar dados na Web é chamado navegador Web, cliente Web ou *Browser*. Projetada inicialmente para comunicação de material textual e gráficos simples, atualmente a Web permite a utilização desde imagens a som, vídeos, filmes interativos, aplicativos e muito mais.

Para que seja possível oferecer este tipo de serviço, é preciso que o navegador Web do usuário, ao acessar o seu endereço, receba como resposta um conteúdo em formato Web. Para tanto, é necessário o desenvolvimento de uma estrutura composta por equipamento e programa que passam a ser os responsáveis pela resposta para as solicitações do conteúdo que se deseja disponibilizar.

O aplicativo a ser instalado e configurado será o servidor Web e receberá e responderá às solicitações dos usuários. Entre as várias opções de aplicativos para atender a esta tarefa, pode ser citado o *Internet Information Services* – IIS da Microsoft. Porém, o mais utilizado tem sido o Apache que, vale lembrar, é mais utilizado em sistemas operacionais Linux, am-

biente para o qual foi desenvolvido. Após a instalação do Apache, deverão ser configuradas as variáveis ambientais e de funcionamento, tais como os endereços, os serviços e as interfaces com outros aplicativos.

Um sistema operacional é um *software* responsável pela interpretação dos comandos, pela criação do ambiente de trabalho, e que consiste na camada intermediária entre o *software* aplicativo e o *hardware*. É o interpretador básico de comandos, e é a interface pela qual o usuário tem acesso aos recursos que o *hardware* oferece. Os exemplos mais comuns de sistema operacional são o Windows – Microsoft e o Linux

A performance do *site* vai estar ligada, entre outros fatores, à velocidade de resposta às solicitações geradas pelos usuários via Web. Para que esta velocidade de resposta seja otimizada, é altamente recomendável que o equipamento que irá processar estas informações seja adequadamente configurado.

Cada uma das tarefas do site pode ser atribuída a um equipamento que irá funcionar para aquele fim, sendo definido como servidor, ou como definem Baldam, Valle e Cavalcanti (2002, p.156):

> Computadores que são normalmente a peça central da rede e gerenciam o fluxo de dados e operações de compartilhamento de *hardware*. Os sistemas hoje em dia podem utilizar mais de um servidor simultaneamente, dividindo os serviços para que se possa ter uma rede mais veloz.

Outros servidores podem ser configurados para atender às demandas de serviços de arquivo, de segurança, de imagem e de aplicação.

Entre os serviços que deverão ser mantidos, destacam-se:

– Servidor de Banco de dados: que compreende os aplicativos que compõem o Sistema Gerenciador de Banco de Dados, dentre eles destacamos o MySQL que vem sendo utilizado em função da facilidade de instalação, por estar disponível para vários ambientes, por não requerer muito do servidor, não ter custo (disponível gratuitamente no *site* http://www.mysql.com/), e por possuir uma performance aceitável. Já o PostgreSQL é um sofisticado sistema gerenciador de banco de dados objeto-relacional, que suporta muitos dos recursos SQL, incluindo controle de transações e funções e tipos definidos pelos

usuários. É considerado uma das melhores opções de banco de dados de código aberto, sendo disponibilizada também uma versão comercial. Mais detalhes podem ser encontrados em: http://www.postgresql.org.br/

– Servidor Web: responsável direto pelas respostas às solicitações que são recebidas pelo *site*, originadas pelo acesso realizado pelos usuários através de seus navegadores. São muitas as opções de navegadores, como Mozilla, Opera, Nestscape e Firebird. Porém, atualmente o domínio do aplicativo da Microsoft é incontestável, sendo portanto o Internet Explorer considerado o padrão de mercado brasileiro.

– Manutenção dos objetos digitais que podem estar armazenados por outras bibliotecas digitais ou mesmo em *sites* com outros fins. Cabe salientar a importância da análise da confiabilidade do material que está sendo acessado, bem como de questões ligadas à elaboração do conjunto de metadados de cada objeto e da preservação deste conteúdo, que estará fora do controle da biblioteca digital que está sendo construída. Se a opção for permitir o auto-arquivamento dos recursos digitais por parte dos usuários identificados, deverá constar do projeto, também, uma interface que permita a inclusão de objetos digitais, bem como de seu respectivo conjunto de metadados.

– Os objetos/obras digitais podem ser desenvolvidos em formato PDF, que permite a elaboração de cópias idênticas aos documentos originais ou com um formato de apresentação inalterável independentes da plataforma computacional, e com um grau de segurança aceitável. Para a criação de textos em formato *Portable Document Format* (PDF) pode ser utilizado o PDF995, que além de rápido e de fácil utilização, não tem custo. Pode ser encontrado em: http://www.pdf995.com/. O aplicativo Adobe Reader da empresa Adobe System Corporation permite o acesso a esses documentos, sem custo, e o *donwload* do aplicativo de acesso e leitura pode ser feito via endereço: http://www.adobe.com/products/acrobat/readstep2.html. Os recursos digitais podem ainda ser apresentados com forma de arquivos compactados. A compactação é a operação que consiste em retirar a redundância dos dados, de forma a

diminuir o tamanho do arquivo, sem perdas de dados. Arquivos compactados podem facilitar o armazenamento e a transmissão, porém o cliente deverá possuir o aplicativo que permite o acesso e a descompactação destes arquivos. Exemplos de programas para compactação de dados: Brazip; WinZip, Winrar e Pkzip – este último utilitário pode ser copiado no *site*: http://www.pkware.com/

– Ferramentas de busca: utilização de uma aplicação de desenvolvimento proprietário exclusivo ou desenvolvida por terceiros, pelas empresas que fornecem gratuitamente ferramentas de busca de uso geral na Internet e que podem ser configuradas para atuação interna e externa ao banco de metadados da biblioteca digital. Segundo Vidotti (2001, p.48),

> As ferramentas de busca, conhecidas também como máquinas de busca ou *search engines*, utilizam *softwares* que indexam e catalogam as páginas e/ou *sites* da Internet em bases de dados, com a finalidade de possibilitar a recuperação de documentos solicitados pelos usuários da Internet, segundo as estratégias de busca e os critérios adotados.

– *Software* de gestão de bibliotecas digitais: existem inúmeros *sofwares* disponíveis para o desenvolvimento de uma biblioteca digital, que vão desde aplicativos para o gerenciamento de coleções digitais até a gestão completa de todas as atividades e serviços de uma biblioteca digital. Esses *softwares* podem ser comercializados ou distribuídos conforme conceito de *softwares* gratuitos e/ou livres. No *software* livre os usuários podem utilizar, estudar, modificar e até mesmo repassar sem que seja necessária autorização do autor.

– Porém, o desenvolvimento do próprio *software* requer uma análise cuidadosa de seus custos em função do retorno a ser obtido e também da questão da continuidade do projeto. Uma solução proprietária é aquela em que a própria organização realiza o desenvolvimento, seja com mão-de-obra própria, seja terceirizada. Entre os pontos positivos de uma solução proprietária, podemos destacar que a solução é construída de acordo com as necessidades identificadas pela organização, além de haver absoluto controle da continuidade do desenvolvimento pela própria organização. Vantagem importante, também, é a

de manter a organização livre de custos de licenças e independente de restrições geradas por um possível fornecedor de aplicativos. Por outro lado, existem pontos negativos a serem considerados, como a necessidade de manutenção de recursos humanos e esforços para documentação do sistema e manutenção do aplicativo e a dificuldade de acompanhamento da evolução da tecnologia. Também é importante levar em consideração a dificuldade de se manter alinhado aos padrões de mercado e de comunicação com outras organizações que os utilizam.

– Vale destacar que a utilização de um *software* para o desenvolvimento de uma biblioteca digital, implica necessariamente a implementação de uma interface de comunicação via Internet, ou seja, a implantação de um *website*, que pode ser desenvolvido utilizando as linguagens Java, PERL, PHP ou simplesmente HTML, para que os usuários possam buscar e visualizar os objetos digitais, contendo, no mínimo, uma ferramenta de busca que permita a localização de um determinado objeto digital ou mesmo a visualização dos objetos com definições aproximadas das solicitadas pelo usuário. Este aplicativo deverá ter acesso à base de metadados da biblioteca digital.

– Preservação: considerando a preservação do documento original, a conversão do documento para a meio digital implica, segundo Levacov (1997, p.5),

> [...] tão somente a preservação da informação e não a preservação física do documento na mídia original. O suporte digital [...] cria novos problemas: o da obsolescência das tecnologias de preservação, armazenamento e recuperação *(hardware, software*, sintaxe etc.) e, dada a facilidade de manipulação de dados pela mídia digital, o da autenticidade. [Enquanto que] o problema da preservação "intelectual" se refere à integralidade e autenticidade dos documentos que podem ser corrompidos intencional ou acidentalmente.

Para que o ambiente informacional não seja alterado de forma indevida, são utilizados *firewalls*, barreiras interpostas entre a rede interna e a rede externa com a finalidade de evitar acessos não autorizados, protegendo programas e equipamentos de ações com origem externa (Internet). São

aplicativos preferencialmente instalados em computadores dedicados somente a esta função. Além disso, devem ser instalados antivírus eficazes que inibam a execução de vírus que se instalam nos equipamentos sem autorização (consciente) do usuário e que podem provocar danos tanto aos dados, como aos aplicativos e ao próprio *hardware*. Manter aplicativos que possam controlar e impedir que esses programas sejam instalados é fator fundamental na política de segurança de qualquer projeto ligado à tecnologia da informação, não sendo diferente no caso das bibliotecas digitais.

4. Considerações Finais

Diante do exposto, podemos afirmar que para a organização de informações que serão disponibilizadas em meios digitais, sua forma de armazenamento, recuperação e apresentação, podem ser utilizados os princípios básicos da Arquitetura da Informação de *website*.

Os elementos básicos da Arquitetura da Informação aplicados à biblioteca digital visam à criação de estruturas digitais que priorizem a organização descritiva, temática, representacional, visual e navegacional de informações, em consonância com o conteúdo informacional, o contexto e o usuário, tendo objetivos bem definidos, adequando o dimensionamento e o direcionamento dos serviços e dos produtos informacionais aos usuários potenciais.

Cunha (1999, p.265) afirma que

> As bibliotecas, em todas as épocas, sempre foram dependentes de tecnologia da informação. A passagem dos textos manuscritos para a utilização de textos impressos, a utilização pioneira do microfilme como agilizador do intercâmbio de informações, o acesso à base de dados bibliográficos armazenados nos grandes bancos de dados e o uso do *cd-rom* demonstram que nossas instituições sempre acompanharam e venceram os novos paradigmas tecnológicos. O advento da biblioteca digital, altamente dependente das diversas tecnologias de informação, também ultrapassará essas novas barreiras.

As bibliotecas digitais, além das atividades oferecidas pelas bibliotecas tradicionais, têm características próprias, que possibilitam a

otimização do uso das tecnologias da informação, agregando valores aos serviços oferecidos, possibilitando ao usuário o acesso independente de tempo e espaço, e os recursos/obras digitais podem conter textos, imagens estáticas e dinâmicas e sons, com características multimídia (auditivas e visuais) e interligadas por meio de *hiperlinks* contextuais. Além disso, o serviço de referência digital pode contar com alguns mecanismos de avaliação do usuário, via interação com o ambiente informacional, de modo a prover automaticamente tanto a disseminação seletiva da informação como a interface comunicacional eficientes e personalizadas.

Nota

[1] A digital library remains a library, with the same purposes, functions and goals as a traditional library. The digital part of the term indicates merely that the material is stored and accessed digitally. A digital library is therefore far more than a digital collection, particularly a collection, as is illustrated in the overwhelming number of the World Wide Web sites on the Internet, which consists of relatively volatile current information.

Referências

AGNER, L. Arquitetura de informação: testes de usabilidade.*Web Insider*. 06 jan. 2004. Disponível em: http://www.webinsider.com.br/vernoticia.php?id=2018 > Acesso em: 30 jan. 2004.

BALDAM, R. de L.; VALLE, R. de A. B.; CAVALCANTI, M. C. B. *GED:* Gerenciamento Eletrônico de Documentos. São Paulo: Érica, 2002. 205p

CAMARGO, L. de A. *Arquitetura da informação para biblioteca digital personalizável*. 2004. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências. Universidade Estadual Paulista, Marília.

CUNHA, M. B. da.Desafios na construção de uma biblioteca digital. *Ciência da Informação*, Brasília, v.28, n.3, p.255-266, set./dez. 1999.

DIAS, C. *Usabilidade na web*: criando portais mais acessíveis. Rio de Janeiro: Alta Books, c2003.

DONATI, L.P.; CARVALHO, H.; PRADO, G. *Sites na Web:* Considerações sobre o design gráfico e estrutura de navegação. Projeto Wawrwt, 1998. Disponível em: <http://www.wawrwt.iar.unicamp.br/entrada_textos01.html>. Acesso em: 03 mar. 2004.

GARRETTT, Jesse James. *The Elements of user experience*: user-centers design for the Web. 2.ed. Indianápolis (Indiana), 2003. 189 p. Disponível em: <http://www.jjg.net/ia/elements.pdf>. Acesso em: 20 mai.2004.

LEVACOV, Marília. Biblioteca Virtuais:(R)evolução? *Ciência da Informação*, Brasília, v.26, n.2, p.125-135, maio/ago. 1997.

McGEE, J.; PRUSAK, L. *Gerenciamento estratégico da informação*. Trad. Astrid Beatriz de Figueiredo. Rio de Janeiro: Campus, p.129-149, 1994.

NUNES, F. O. *A arquitetura de navegação do ciberespaço*. Web Arte no Brasil. 2000. Disponível em: < http://webartenobrasil.vilabol.uol.com.br/artigo6.htm > Acesso em: 10 mar. 2004.

ROSENFELD, L.; MORVILLE, P. *Information architecture for the World Wide Web*. Sebastopol, CA: O'Reilly, 1998. 202p.

SARMENTO E SOUZA, M. F.; FORESTI, M.C.P.P.; VIDOTTI, S.A.B.G. Ar-

quitetura da informação em web site de periódico científico. *ETD – Educação Temática Digital*, v.5, n.2, 2004. Disponível em: <http://143.106.58.55/revista/viewarticle.php?id=13&layout=abstract>. Acesso em 10 mar. 2005.

SIQUEIRA, M.A.; SANTOS, P.L.V.A. da C. A versão em XML do MARC21 e as formas de representação descritiva na Ciência da Informação. In: VIDOTTI, S.A.B.G. (Coord.). Tecnologia e conteúdos informacionais. São Paulo: Polis, 2004. p.95-111

STRAIOTO, F. *A Arquitetura da informação para a World Wide Web:* um estudo exploratório. 2002. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências. Universidade Estadual Paulista, Marília.

VIDOTTI, S.A.B.G. *O Ambiente hipermídia no processo de ensino-aprendizagem*, 2001. 125 f. Tese (Doutorado em Educação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – UNESP, Marília.

WAINWRIGHT, E. *Digital libraries:* some implications for government and education from the Australian development experience. Canberra, Austrália: Nacional Library of Austrália, 1996. Disponível em: <http://www.nla.gov.au/nla/staffpaper/ew6.html>. Acesso em: 10 mar. 2005.

Metadados: descrição e recuperação de informações na Web

*Carlos Henrique Marcondes**

1. Introdução

Na história das bibliotecas, à medida que as coleções aumentam, descrever documentos vem se tornando uma condição fundamental para possibilitar sua posterior recuperação entre os itens da coleção, avaliação de relevância e posterior uso. Profissionais de informação vêm criando, há séculos, metodologias para registro, inventário e descrição de documentos, como forma de controlar acervos e prover meios de acessar seletivamente os itens de uma coleção. Um instrumento de pesquisa de um arquivo ou um catálogo de uma biblioteca nada mais são que descrições de documentos de uma coleção, organizadas com a finalidade de facilitar sua recuperação e acesso, os agora chamados metadados (Milsted, 1999).

O que antes era um problema dos profissionais de informação, em contextos específicos como a recuperação de informação em arquivos e bibliotecas, generaliza-se com o surgimento da Web em meados dos anos 90. Parcelas significativas da cultura contemporânea, entretenimento, ciência, educação e negócios passam a estar organizados em torno da Web.

* Professor do Departamento de Ciência da Informação, UFF, marcon@vm.uff.br

O termo "metadados" surge neste contexto: o crescimento acelerado e caótico da Web coloca na ordem do dia o problema de identificar, recuperar e avaliar a infinidade de recursos os mais diversos, tornados disponíveis na Web.

Um dos maiores objetivos do uso de metadados no contexto da Web é permitir não só descrever documentos eletrônicos e informações em geral, possibilitando sua avaliação de relevância por usuários humanos, mas também permitir agenciar computadores e programas especiais, robôs e agentes de *software*, para que eles compreendam os metadados associados a documentos e possam então recuperá-los, avaliar sua relevância e manipulá-los com mais eficiência. Nesta questão é que entra a linguagem XML – Extensible Markup Language – Linguagem de Marcação Extensível – padrão proposto e homologado pelo W3C[1], como mecanismo de codificar metadados associados a um documento eletrônico em um formato que seja legível não só por pessoas *mas também por programas*. O uso de programas para processar metadados em XML é um dos pilares da iniciativa chamada Web semântica (Berners-Lee, 2001).

A identificação e recuperação de recursos informacionais torna-se assim uma das questões mais importantes da atual economia da Web. Como já dizíamos (Marcondes, 2001), "... a informação relevante para um dado problema tem que estar disponível no tempo certo. De nada adianta a informação existir se quem dela necessita não sabe da sua existência ou se ela não puder ser encontrada."

A literatura brasileira é ainda escassa e esparsa sobre o assunto (Souza, 1997), (Rosseto, 2002). As experiências práticas brasileiras mais conhecidas se referem ao desenvolvimento do Padrão Brasileiro de Metadados de Teses e Dissertações Digitais, o MTbr (Marcondes, 2003), da Biblioteca Brasileira de Teses e Dissertações (BDTD) operada pelo IBICT. Outra experiência significativa é a implementação do Open Archives Protocol for Metadata Harvesting no SciELO (Marcondes, 2003b).

Este capítulo discute estas atividades de descrição de informações – o uso dos assim chamados metadados – no contexto da Web e das publicações eletrônicas e sua importância para a recuperação e a economia da informação. Embora o conceito de metadados não seja limitado ao uso na descrição e recuperação de informações, apresenta e discute com mais detalhes o padrão Dublin Core, o mais conhecido para esta finalidade.

2. Web, metadados e descoberta de recursos

Metadados são definidos como "dados sobre dados" (Weibel, 1995). São dados associados a um recurso Web, um documento eletrônico, por exemplo, que permitem recuperá-lo, descrevê-lo e avaliar sua relevância, manipulá-lo (o tamanho de um documento, ao se fazer *downloading* ou o seu formato, para sabermos se dispomos do programa adequado para manipulá-lo), gerenciá-lo, utilizá-lo, enfim.

A primeira tentativa de dar conta da explosão informacional em que se transformou a Web foram os catálogos, como o Yahoo (o primeiro catálogo da Web), e os chamados mecanismos de busca, como AltaVista, Lycos, WebCrawler etc, e mais recentemente, o Google. Enquanto em catálogos como o Yahoo, a descoberta, avaliação e descrição e inclusão dos recursos Web na base de dados são feitas por profissionais de informação, os mecanismos de busca, para indexarem a Web, possuem programas que visitam página por página da Web, percorrem o texto de cada página, extraindo daí palavras-chaves, e armazenando numa base de dados estas palavras-chaves, associadas ao URL da página. É sobre esta base de dados que os usuários fazem suas buscas nos *sites* dos mecanismos de busca. Naturalmente, por ser uma indexação automática com base em palavras isoladas, sem nenhum controle terminológico, efetuada em páginas sobre os mais variados assuntos, diferentes idiomas e totalmente desprovida de qualquer informação contextual, os resultados têm baixíssima precisão (Sneiderman, 1997). Estudos como o citado reforçam a dimensão do problema localização/identificação colocado pela Internet.

Com o objetivo de ajudar a obter maior precisão nas buscas por páginas Web, ajudando os robôs dos mecanismos de busca a fazerem uma indexação de maior qualidade, num primeiro momento foram incorporados metadados no texto mesmo destas páginas. Isto foi feito com o uso de *tags* especiais da linguagem HTML (a linguagem em que são escritas as páginas da Web), as *tags* META, como mostrado a seguir:

```
<meta name = "cd.subject" content = "bibliotecas digitais, metadados;
recuperação de informações; XML">;
```

Ou então:

```
<meta name = "dc.creator" content = "MARCONDES, Carlos Henrique">.
```

Estas *tags* não ficam visíveis quando a página HTML é exibida na tela de um *browser* como Internet Explorer ou Nestcape Navigator, mas podem ser processadas por um programa robô que esteja indexando a página.

No entanto, esta tentativa inicial esbarra com um outro problema, a emergência da chamada Web profunda ou *deep Web* (Bergman, 2001). Após um período inicial da Web, baseado em páginas estáticas escritas na linguagem HTML e armazenadas como arquivos .html ou .htm nos diferentes servidores Web – a chamada Web superficial –, cada vez mais documentos eletrônicos só se tornam acessíveis através de interfaces de bancos de dados, invisíveis portanto aos programas-robôs dos mecanismos de busca. Esta é a chamada Web profunda. Nestes casos, as páginas HTML são geradas pelos sistemas de bancos de dados *on-the-fly*, ou seja, dinamicamente, como resultado de uma consulta específica, só para aquele usuário que fez a consulta, e não mais como arquivos armazenados permanentemente no disco rígido de um servidor. Segundo este autor, a Web profunda é de 1000 a 2000 vezes maior que a Web superficial.

Exemplos desta situação podem ser vistos quando consultamos o *site* da Receita Federal para saber se a devolução do Imposto de Renda já foi depositada em nossa conta corrente. Quando informamos o CPF, o sistema de banco de dados emite automaticamente um documento em HTML só para nós. Caso semelhante acontece quando consultamos o *site* do IBGE, cruzando mortalidade infantil com determinado município, com determinado ano. Em casos como os citados, esquemas como as *tags* META, que são parte de um documento HTML armazenado num servidor, não funcionam, porque os programas-robôs dos mecanismos de busca, ao percorrerem determinado *site*, não encontrarão estas páginas.

O problema da Web profunda faz com que, hoje em dia, cada vez mais metadados esteja armazenados em bancos de dados ou catálogos de bibliotecas digitais, associados, através de URL, aos documentos eletrônicos que eles descrevem. Conjuntos de metadados relativos a um documento eletrônico também não precisam ser únicos: vários conjuntos de metadados, produzidos por diferentes instituições, podem ser as-

sociados ao mesmo documento eletrônico. Desta forma, o mesmo documento eletrônico pode estar sendo referenciado por vários conjuntos de metadados, assim como um livro pode ter cópias em diversas bibliotecas e em cada uma delas ser descrito diferentemente, com várias finalidades diferentes, não só para descrevê-lo e ajudar na sua recuperação, como também para ajudar a sua gestão, identificar questões relativas a direitos autorais ou à preservação digital do documento.

A descoberta de recursos para viabilizar seu uso torna-se, como já foi dito, altamente crítica num ambiente como a Web. Diferentemente do contexto restrito dos catálogos de uma biblioteca, a Web é um ambiente compartilhado mas altamente desestruturado, gerido de forma descentralizada e em crescimento explosivo, constituindo um verdadeiro desafio para seus gestores, em especial para os profissionais da informação. Todas as ações e decisões dos profissionais de informação têm um impacto global, tudo que for publicado é imediatamente visível para o mundo todo. Estes têm de ter a consciência que este desafio só pode ser enfrentado com estratégias que, embora já conhecidas pelos profissionais de informação, agora têm que ser empregadas numa escala global. Estamos falando de cooperação e compartilhamento de recursos.

No espaço da Internet, atividades como a descoberta de recursos são desenvolvidas não só por usuários humanos, mas também, e cada vez mais, na perspectiva da Web semântica, por meios automáticos, por programas robôs, "aranhas", agentes de *software*. Isto é motivado pelo grande crescimento da Web e esta é uma das principais estratégias de fazer frente a este crescimento explosivo.

Além do uso de agentes de *software*, outra estratégia igualmente importante, ligada também à questão dos metadados, é, uma vez que as publicações na Web crescem de forma exponencial e não existem profissionais de informações suficientes para descrevê-los todos, *permitir que autores eles mesmos descrevam seus documentos ao publicá-los eletronicamente* "na fonte", conforme uma proposta bem antiga da biblioteconomia, retomada por Weibel (1995).

Para isso é necessário portanto que descrições/representações de recursos informacionais – os agora chamados metadados – possam seguir um padrão de descrição que *seja simples o suficiente para que um autor possa descrever seu documento*. Esta é a proposta da iniciativa Dublin Core.

3. O padrão Dublin Core e sua codificação legível por máquina

O padrão de metadados para descrição de documentos eletrônicos, proposto pela Dublin Core Metadata Initiative – DCMI (http://purl.org/dc), foi pensado segundo esta idéia: ser simples o suficiente e auto-explicativo, de modo que o próprio autor do documento possa ele mesmo, descrevê-lo ao publicá-lo eletrônicamente. Outros padrões de descrição/representação de documentos, como o MARC por exemplo, são mais complexos, exigindo o aporte de profissionais de informação treinados.

O padrão DCMI é composto de somente 15 elementos descritivos, simples e auto-exeplicativos, o chamado Dublin Core Metadata Element Set – DCMES. Nos sistemas de publicação eletrônica, ao submeter seu documento numa biblioteca digital ou num arquivo de *pre-prints*, o autor, ao incluir ou fazer *upload* do texto, preenche um formulário descrevendo seu documento, associando metadados ao mesmo e armazenado-os num catálogo automatizado ou base de dados.

O conjunto de metadados DCMI e sua utilização são os seguintes:

- Subject (assunto);
- Title (título);
- Creator (autor, responsável pelo documento);
- Publisher (publicador, quem torna o documento disponível na Internet);
- Contribuitor (outros colaboradores - por ex. editores, tradutores, etc);
- Description (descrição, resumo, sumário);
- Date (data de publicação);
- Type (tipo de recurso – homepage, romance, poesia, *software*, dicionário);
- Format (formato do arquivo que contém o documento eletrônico – texto, PDF, LaTex, HTML, WORD ou outro);
- Indetifier (geralmente o URL de um documento eletrônico);
- Relation (relacionamentos com outros documentos, por ex. versões);
- Source (fonte ou origem);
- Language (idioma do documento);
- Coverage (cobertura – espacial ou temporal, sobre que lugar, ou sobre que época o texto se refere);

• Rights (texto livre especificando qualquer restrição referente a direitos autorais).

A concisão da descrição de um documento obtida pelo uso dos 15 elementos descritivos Dublin Core vem sendo objeto de discussão entre os participantes da DCMI. Esta concisão resulta em perda de precisão na recuperação de informações. Para superar estes problema, a DCMI propôs também os chamados *Qualificadores* (http://purl.org/dc/documents .dcmes-qualifiers-20000711.htm) de alguns dos elementos do conjunto, por exemplo (utilizando-se aqui também as tags META da linguagem HTML para codificação):

```
<meta name="dc.Subject scheme="DECS" content="Sida"> ou <meta
name="dc.Subject scheme="CNPq" content="Ciências Sociais Aplicadas">.
```

Nestes casos, fica explícito, através do qualificador *scheme*, que o assunto (dc.Subject) foi retirado de um vocabulário controlado ou de uma tabela de assuntos padronizada (um *scheme*), como o DECS – Descritores em Ciência da Saúde – ou a tabela de áreas de conhecimento CNPq; ou então

```
<meta name="dc.Coverage.temporal" content="Idade Média">.
```

Aqui, o elemento Dublin Core (dc) *Coverage* recebe um "refinamento", especificando que a cobertura é "temporal".

Opositores desta tendência dentro da DCMI argumentam que o uso de Qualificadores faria com que o conjunto de metadados DC perdesse seu caráter intuitivo, dificultando seu uso pelos próprios autores. No entanto, o uso de qualificadores, bem como de qualquer dos 15 elementos, é opcional: pode-se utilizar todos os 15 elementos, somente alguns deles, com ou sem qualificadores.

Uma vez visto o padrão proposto pela DCMI, fica ainda a questão de como codificá-los em meio legível por computador, de forma ampla, generalizada e não exclusiva, para que os metadados possam ser lidos por programas. A forma que tem se firmado como um padrão na Web para codificar metadados tem sido a linguagem XML.

4. Codificando metadados em linguagem XML

O uso das tags META da linguagem HTML para a codificação de metadados, com as suas limitações, foi discutido nos exemplos anteriores. A linguagem XML vem sendo usada de forma cada vez mais generalizada, com esta finalidade. XML é uma linguagem cuja proposta é ser um padrão universal, referendado pelo W3C (http://www.w3c.org/XML/) e aberto, voltado para a *descrição da estrutura de documentos eletrônicos* onde o conteúdo e sua descrição compõem um único arquivo, de tipo texto puro – portável, portanto sem problemas para qualquer tipo de sistema computador, já que este formato, também conhecido como ASCII – American Standard Code for Information Interchange –, é o mais universal e pode ser lido por todas as plataformas computacionais.

Ao contrário de outras formatos de arquivos, os chamados formatos proprietários como o formato .DOC do Microsoft Word ou o formato .MDB do Microsoft Access, um arquivo XML não está preso a um programa para lê-lo; na XML são explicitados, além dos conteúdos dos documentos, a sua estrutura, sob a forma de marcações, identificadas por um nome significativos, envolvido nos caracteres "<" e ">"; a cada marcação indicando o início de um determinado conteúdo, por exemplo <autor>, corresponde uma marcação de fim, </autor>; os conteúdos ficam entre marcações, <autor>João da Silva</autor>, assim:

```
<referencia idioma="pt">
    <autor>Décio, Otávio C</autor>
    <titulo>XML: guia de consulta rápido</titulo>
    <imprenta>
        <local>São Paulo</local>
        <editor>Novatec</editor>
        <data>2000</data>
    <imprenta>
</referencia>
```

Todo documento XML é dito "bem formado" quando pode ser associado a uma árvore, que deve ter uma marcação raiz (no caso, <referencia, </referencia>); a toda marcação inicio deve corresponder uma marcação fim (<autor> e </autor>, <titulo> e </titulo>, etc)

As linguagens para marcação de textos tiveram início com a SGML – Structured Markup Language – e têm na HTML – Hypertext Markup

Language – que se constitui na base da "teia global" da Internet, sua versão mais conhecida. A diferença entre a HTML e a XML é que naquela, as *tags* são *predefinidas*, *imutáveis*; por exemplo, <h1> e </h1> para cabeçalhos, <a href> e para hiperlinks, etc. Todo documento deve ser identificado como HTML (<html> </html>), ter uma área de cabeçalho (<head></head>) com o nome para o documento (<title> </title>), um título principal e uma área definida como corpo (<body></body>) do conteúdo do documento. Como o exemplo a seguir:

```
<html>
<head>
<title>Exemplo de HTML simples</title>
</head>
<body>
<h1>Este é o primeiro nível de cabeçalho</h1>
Bem-vindo ao mundo do HTML.
Este é o primeiro parágrafo.<p>
E este é o segundo.<p>
</body>
</html>
```

Esta página vai aparecer assim na tela de um *browser*:

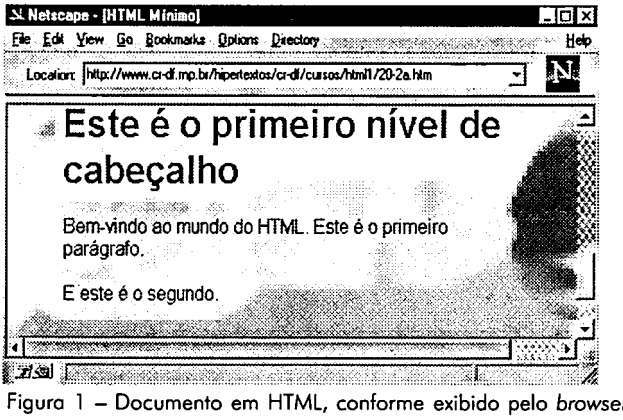


Figura 1 – Documento em HTML, conforme exibido pelo *browser*

Num documento XML as marcações *não têm a ver* com a aparência do documento ao ser exibido num *browser* e sim *com sua estrutura*. O ambiente eletrônico de funcionamento de um documento XML na Web é bastante complexo. Na verdade, um documento XML pode fazer referência a

uma folha de estilos – CSS, *cascating style sheet* –, com instruções referentes especificamente a como o documento deve ser exibido. Um documento XML também pode fazer referência a um outro arquivo, chamado esquema (XMLSchema) ou DTD (Document Type Definition) que especifica regra de validade do documento, por exemplo, quantas ocorrências do campo <autor> são possíveis, que conteúdo é válido para o campo - letras, números, valores numéricos. O uso de documentos XML associados a esquemas ou DTDs garante uma grande precisão na sua formatação, permitindo que eles possam ser usados em transações "seguras" via Internet, como, por exemplo, em comércio eletrônico. Para evitar confusões semânticas sobre os campos usados em documentos XML, o uso do campo <autor> num livro ou numa ação judicial, documentos XML também podem fazer referência aos chamados espaços de nomes – *name spaces*, vocabulários específicos que permitem identificar, através de uma URL, em que contextos e que comunidades utilizam um determinado significado para os campos de documentos XML. Os elementos que compõem o ambiente eletrônico de funcionamento de um documento XML na Web podem ser vistos de forma integrada na seguinte ilustração:

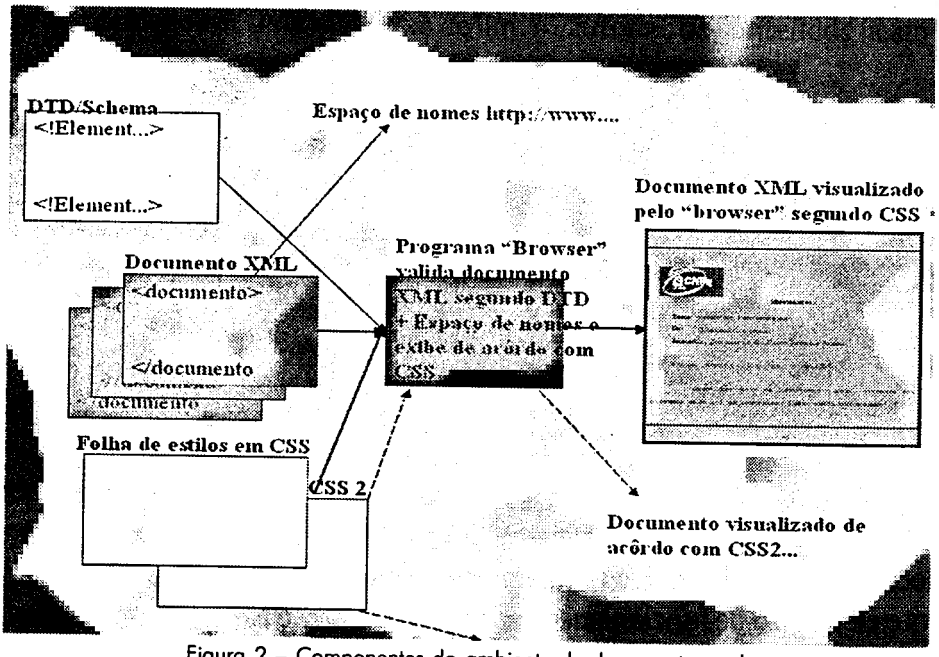


Figura 2 – Componentes do ambiente de documentos xml

Um exemplo do uso de metadados segundo o padrão Dublin Core e codificados em XML é o protocolo Open Archives Initiative Protocol for Metadata Harvesting – OAI-PMH (http://wwwopenarchives.org/OAI/openarchivesprotocol.htm). Este protocolo surge da comunidade de científica que publica seus trabalhos em arquivos eletrônicos de acesso aberto, os *open archives* (http://www.openarchives.org), que começam a surgir como alternativa aos periódicos publicados pelos grandes editores em fins da década de 90, para publicação direta pelos próprios autores (Marcondes, 2001). O OAI-PMH permite a coleta automática de metadados de documentos armazenados em arquivos de publicações eletrônicas – os "provedores de dados". Uma vez coletados, metadados de documentos eletrônicos de diferentes provedores de dados podem ser "reutilizados", por exemplo, agregados numa base de dados única, permitindo realizar uma busca unificada a esta base como se tivesse sendo feita aos diferentes provedores de dados simultaneamente. Os metadados são coletados segundo o padrão Dublin Core. A seguir é mostrado o *diálogo entre os programas: harvest do provedor de serviços e servidor OAI-PMH*; o programa *harvest* solicita metadados de determinado documento, recebe como respostas estes metadados, segundo o padrão Dublin Core e codificados em XML:

- solicitação, pelo programa *harvest*, de metadados segundo o padrão Dublin Core ("oai_dc"), do documento identificado por "oai:arXiv:cs/0112017", do arquivo de *pre-prints* "arXiv.org", enviada segundo um comando HTTP:

http://arXiv.org/oai2?verb=GetRecord&identifier=oai:arXiv:cs/0112017&metadataPrefix=oai_dc

```
<metadata>
<oai-dc:dc xmlns:oai-dc="http://www.openarchives.org/
xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>A reassessment of the role of serotonergic system in the control of
behavior</dc:title>
<dc:creator><Medeiros,Magda A.></dc:creator>
<dc:creator><Costa-e-Sousa,Ricardo H.></dc:creator>
<dc:creator><Olivares,Emerson L.></dc:creator>
<dc:creator><Cortes,Wellington S.></dc:creator>
<dc:creator><Reis,Luis C.></dc:creator>
<dc:subject><food ingestion></dc:subject>
<dc:subject><serotonergic system></dc:subject>
<dc:subject><obesity></dc:subject>
<dc:description><The role of serotonergic system in the feeding ... </dc:description>
<dc:publisher><Anais da Academia Brasileira de Ciências></dc:publisher>
<dc:date>2005-03-01</dc:date>
<dc:type>journalarticle</dc:type>
<dc:format>text/html</dc:format>
<dc:identifier>http://www.scielo.br/scielo.php?script=sci_
37652005000100008</dc:identifier>
<dc:language>en</dc:language>
</oai-dc:dc>
/metadata>
```

Este documento usa campos referentes a dois "names spaces" ou vocabulários diferentes, Open Archives Initiative e Dublin Core

Ele aponta também para o seu esquema, o

Os elementos to especo de nomes "Dublin Core" são prefixados por "dc:", como <dc:title>, <dc:creator>, etc

Este é o endereço eletrônico do documento, permitindo, através de um "click" acessa-lo.

Uma aplicação especial em XML para a descrição de recursos Web é o RDF –Resource Description Framework (http://www.w3c.org/RDF/), também um padrão W3C. Enquanto XML é uma linguagem genérica para a descrição da estruturação de documentos eletrônicos, RDF é específica para criar metadados com a finalidade de localizar e identificar recursos. RDF usa XML dentro de um esquema mais estruturado. Baseia-se na concepção de que um recurso Web possui propriedades (por exemplo, seu autor, seu título, seu idioma) e que cada uma destas propriedades possui um valor ("João da Silva", "Minha vida", "português"). O valor de uma propriedade pode ser outro recurso, por exemplo, o valor de autor pode ser o endereço de sua página Web. Um exemplo é mostrado a seguir

```
<?xml version="1.0"?>
<rdf:RDF
        xmlns:rdf="http://wwww3.org/1999/02/22-rdf-syntax-ns"
        xmlns:cd="http://www.recshop.fake/cd"
                <rdf:Description rdf:about="http://www.recshop.fake/cd/Brothers_in
                Arms">
                <cd:artist>Dire Strait</cd:artist>
                <cd:country>UK</cd:country>
                <cd:company>http://www.polygram.com</cd:company>
                <cd:price>10,90</cd:price>
                <cd:year>1987</cd:year>
        </rdf:Description>
</rdf:RDF>
```

A descrição de um recurso em RDF é escrita em XML

O recurso descrito é indicado aqui, pela sua URL

Usa campos referentes a dois "names spaces" ou vocabulários diferentes, RDF e Recshop

tíficos este vocabulário poderia ser o Dublin Core, conforme visto em Beckett (2002).

5. Conclusões

De forma acelerada, mais e mais produtos da cultura humana passam a ser publicados e distribuídos diretamente em meio digital através da Web, pois esta proporciona alcance, rapidez, baixo custo e interatividade. A Web vem se tornando algo como uma única e imensa biblioteca digital. Um recurso informacional, com esta dimensão, é único na história da cultura humana. Nossas responsabilidades, como profissionais de informação, para a gestão e o uso comum destes recursos também assumem uma dimensão mais ampla.

Mais e mais se torna premente facilitar a recuperação destes recursos para que eles possam ser utilizados. A Web vem tendo um crescimento ao mesmo tempo acelerado e caótico, e a ordenação deste crescimento, o uso de padrões que facilitem a recuperação e viabilizem o uso das informações disponíveis tem sido uma preocupação constante. O uso de metadados pelas mais diferentes comunidades da Web se insere neste contexto.

Dublin Core é o mais o mais específico para informações bibliográficas. Outros padrões de metadados bastante usados em outras áreas especializadas são os seguintes:

- Material instructional/pedagógico - IMS – Instruction Management System, http://www.imsproject.org/metadata/index.html;
- Dados geográficos – Content Standard for Digital Geospatial Metadata (CSDGM), http://www.fgdc.gov/metadata/contstan.html
- Metadados de descrição arquivística em formato legível por computador - EAD – Enconding Archives Description, http://locweb.loc.gov/ead/ead.html;
- Informações governamentais - GILS – Global Information Locator Service, http://www.gils.net/elements.html
- Comércio eletrônico - INDECS – Interoperability of Data in E-Commerce Systems, http://www.indecs.org
- Metadados para preservação digital - OAIS – Open Archiving Information System, http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/

documents/pdf/CCSDS-650.0-B-1.pdf

• Metadados sobre direitos autorais – project RoMEO, http://www.lboro.ac.uk/projects/romeo/index.html

Dublin Core é um padrão voltado principalmente para a descrição e a descoberta de documentos eletrônicos. Caso a biblioteca digital, além de referenciar recursos eletrônicos, também seja uma publicadora ou editora, como o são várias bibliotecas digitais especializadas ou arquivos eletrônicos, que têm a custódia dos documentos que compõem sua coleção, terá que se preocupar, além de referenciar seus documentos com metadados segundo padrões como Dublin Core, com questões como *copy right*, administração e gestão destes documentos e a preservação digital dos mesmos. Para todas estas atividades, existem conjuntos de metadados específicos.

XML é, cada vez mais, a base de vários padrões da área de informação. Além do OAI-PMH, outro protocolo de recuperação de informações, o SRW – "Search Retrieval on Web" (http://www.loc.gov/z3950/agency/zing/srw/) -, o sucessor do protocolo Z39.50, de largo uso em bibliotecas para viabilizar a consulta a um catálogo virtual formado por catálogos distribuídos de várias bibliotecas, tem por base para troca de mensagens entre programas clientes e servidores o XML. As ontologias, bases de conhecimento sobre determinado domínio, baseiam-se na OWL – "Ontology Web Language" (http://www.w3c.org/2004/OWL/), codificada em XML.

Os problemas de recuperação de informações, de uma maneira ampla, vêm sendo endereçados pela iniciativa chamada Web semântica (http://www.w3c.org/2001/sw/). A Web semântica (Souza, 2004) é uma evolução da Web atual, baseada em documentos hipertextuais escritos segundo a linguagem HTML, que só permitem a indexação automática por palavras chaves, extraídas do texto. A base da infra-estrutura da Web semântica vão ser páginas escritas em XML, que, além do conteúdo, terão metadados utilizando vocabulários e relações muitos poderosos (utilizando RDF e ontologias) para expressar a semântica das novas páginas Web. Sobre esta estrutura de páginas ricas semanticamente irão atuar programas especiais, os chamados agentes inteligentes, que terão acesso ao conhecimento embutido nas páginas da Web semântica, para rea-

lizarem tarefas que demandem conhecimento, raciocínio, dedução. Espera-se, desta forma, que estes programas possam ser muito mais cooperativos e úteis, facilitando a vida dos usuários humanos, antecipando-se a suas necessidades, *aprendendo*, com o tempo, os hábitos e necessidades de seus usuários. Espera-se com isso caminhar para a realização da proposta do idealizador da Web semântica, Tim Berners-Lee (2001), que nos é muito familiar, lembrando a proposta do Mundaneum de Paul Otlet: "A Web semântica não é meramente a ferramenta para conduzir tarefas individuais que nós temos discutido até aqui. Mais que isso, se adequadamente planejada, A Web semântica pode apoiar a evolução do conhecimento humano como um todo".

Notas

[1] World Wide Web – consórcio internacional que trabalha na proposição de padrões tecnológicas para a Web,

http://www.w3c.org

Referências

BECKETT, David; MILLER, Eric; BRICKLEY, Dan. *Expressing simple dublin core in RDF/XML*. 2002. Disponível em <http://dublincore.org/documents/2002/07/31/dcmes-xml/>. Acesso em 02 dez. 20004.

BERGMAN, Michael K. The deep web: surface hidden value. Journal of Electronic Publishing, v.7, n.1, 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

Dublin Core Qualifiers. Disponível em <http://purl.org/dc/documents/rec/dcmes-qualifiers-20000711.htm>.

Dublin Core metadata elements. ANSI/NISO Z39.85-2001. NISO : 2001. Disponível em <http://www.niso.org>. Acesso em 29 out. 2001.

INTRODUCTION TO RDF. Disponível em <http://www.w3.schools.com/rdf/rdf_intro.asp:output=print>. Acesso em 29 jan. 2005.

MARCONDES, Carlos Henrique. Representação e economia da informação. *Ciência da informação*, Brasília, v.30, n.1, p.61-70, 2001. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000100008&lng=pt&nrm=iso>. Acesso em 05 maio 2002,

MARCONDES, Carlos Henrique, SAYÃO, Luis Fernando. Acesso unificado às teses eletrônicas brasileiras. *Informação sociedade estudos*, João Pessoa, Paraíba, v.13, n.1, 2003. Disponível em <http://www.informacaoesociedade.ufpb.br>

MARCONDES, Carlos Henrique; SAYÃO, Luis Fernando. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. *Ciência da informação*, Brasília, v.31, n. 3, p.42-54, set./dez.2002. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000300005&lng=pt&nrm=iso>. Acesso em 27 fev.2003.

MARCONDES, Carlos Henrique, SAYÃO, Luis Fernando. The SCIELO Brazilian scientific journal gateway and Open Archives: the development of SciELO / OpenArchives Data Provider server. DLib Magazine, 2003. Disponível em <http://www.dlib.org/dlib/march03/marcondes/03marcondes.html>.

MEDEIROS, Norm. XML and the resource description framework: the great web hope. *ON-LINE*, set. 2000. (http://www.onlinemc.com/onlinemag/OL2000/mederios89.html, visitado em 09/01/2001).

MILSTED, Jessica, FELDMAN, Susan. Metadata: cataloging by any other name. *ONLINE*, jan. 1999. (http://www.online.com/onlinemag/ OI1999/milstead1.html, visitado em 28/08/99).

ROSSETO, Márcia. METADADOS: novos modelos para descrever recursos de informação digital. In: INTEGRAR: Congresso Internacional de Arquivos, Bibliotecas, Centros de Documentação e Museus, São Paulo, 17-22 março, 2002, 1, *Anais*... p. 485-498.

SHNEIDERMAN, Ben. Clarifying search: a user-interface framework for text searches. *Dlib Magazine*, January 1997. Disponível em <http://www.dlib.org/dlib/january97/retrieval/01sneiderman.html>. Acesso em 5 maio 2001.

SOUZA, Renato Rocha, ALVARENGA, Lídia. A Web semântica e suas contribuições para a Ciência da Informação. *Ciência da Informação*, Brasília, v.33, n.1,. 2004. Disponível em <http://www.ibict.br/cienciadainformacao/viewarticle.php?id=71&layout=html>. Acesso em 15 out. 2004.

SOUZA, Terezinha Batista de, CATARINO, Maria Elisabete, SANTOS, Paulo Cesar dos. Metadados: catalogando dados na Internet. *TransInformação*, Campinas, v. 9, n. 2, p.93-105, maio/ago. 1997.

USING DUBLIN CORE IN XML. Disponível em <http://purl.org/dc/documents/wd/dcmes-xml20000714.htm>.

WEIBEL, Stuart. Metadata: the foundations of resource description. *D-Lib Magazine*, July 1995. Disponível em <http://www.dlib.org/dlib/July95/07weibel.htm>. Acesso em 15 maio 2001.

Preservação digital no contexto das bibliotecas digitais: uma breve introdução

Luís Fernando Sayão[*]

1. A era do esquecimento

São muitos os efeitos negativos das tecnologias modernas, e os seus desdobramentos – se nada for feito – nos aguardam numa curva do futuro. O mais dramático e o mais discutido é certamente a degradação ambiental, o leviatã que ameaça a própria vida no nosso planeta. Porém, sem ao menos saber, estamos submetidos a vários outros perigos decorrentes do uso inconseqüente da tecnologia. Dentre os muitos problemas provocados pelo avanço vertiginoso da tecnologia – assinalados na agenda crítica da humanidade à espera de uma solução – está o risco da perda irreversível dos estoques de informações digitais, que são cada vez mais os principais registros das atividades, de toda a natureza, da sociedade moderna.

O dilema que se instala é que tecnologia digital coloca a humanidade sob o risco de uma amnésia digital, que já está em curso, ao mesmo tempo em que abre oportunidades extraordinárias em todos os campos – da ciência à arte, do trabalho ao lazer. A urgência da sociedade contemporânea em transformar tudo, tudo – textos, imagens, vídeos, música – para formatos digitais é justificada pela enorme economia de espaço físi-

[*] Comissão Nacional de Energia Nuclear – CNEN. Centro de Informações Nucleares – CIN
lsayao@cnen.gov.br

co de armazenamento e, sobretudo, pelos extraordinários ganhos de produtividade e eficiência proporcionados pela otimização dos fluxos de trabalho. Isto sem falar nas facilidades de acesso aos estoques informacionais, públicos e privados, que, por si só, são um fenômeno novo e sem precedentes, como também é a sua outra face: a facilidade de produção, edição, publicação, integração e distribuição de informação em formatos digitais pelos meandros das redes de computadores mundiais.

Dentro desse contexto de incertezas, talvez não haja desafio maior para os bibliotecários, arquivistas e demais profissionais de informação e conhecimento, neste começo de século, do que garantir meios de acesso à informação digital às futuras gerações, para que elas não sejam ameaçadas ou sofram os efeitos de algo catastrófico como uma era do esquecimento. A questão essencial que se coloca para a sociedade da informação pode bem ser como salvar a nossa memória digital [1].

Não é necessário dizer que grande parte da informação produzida hoje já nasce digital, e as que estão em suportes convencionais estão rapidamente sendo convertidas para formas digitais; o que é preciso enfatizar é que, apesar dessa tendência avassaladora e irreversível, não há em contrapartida estratégias consolidadas para garantir o acesso de longo prazo às informações digitais de valor contínuo. Muitas informações consideradas tesouros digitais pela Unesco[29] já foram irremediavelmente perdidas, como foi, por exemplo, a primeira mensagem eletrônica enviada por cientistas do MIT, Massachusetts Institute of Technology, em 1964 [1]. A nossa compulsão em produzir informações digitais é infinitamente superior à nossa capacidade de preservar o acesso a elas. Não obstante, o que a humanidade deseja, sem talvez dar conta da dimensão do problema, é garantir que a herança cultural, histórica, científica e econômica, expressa através de informação digital, possa no futuro ser acessada por meio dos recursos tecnológicos disponíveis na época, de maneira íntegra, e que essa informação tenha a garantia também de sua autenticidade e confiabilidade – o seu valor de prova.

O maior problema da preservação digital é que a tecnologia digital, em comparação com a impressão tradicional, é um suporte extremamente frágil e instável. A longevidade dos materiais digitais está ameaçada pela vida curta das mídias digitais, pela obsolescência rápida dos equipamentos de informática, dos *softwares* e dos formatos. O tempo cada vez mais curto entre a inovação e a obsolescência tecnológica nas áreas de tecnologia da informação constitui

uma ameaça cada vez mais contundente à longevidade dos objetos do reino digital. Isto acontece principalmente porque a rápida obsolescência é a chave da sobrevivência empresarial num mercado altamente competitivo.

Entretanto, o impacto da curta longevidade dos materiais codificados digitalmente não está circunscrito somente aos sistemas de informações formais – arquivos e bibliotecas digitais –, ou à consciência histórica e à memória coletiva das organizações e dos seres humanos. Ele tem outros desdobramentos, como o ilustrado pela seguinte questão: como poderemos lançar para um futuro distante mensagens suficientemente duradouras, alertando sobre depósitos de rejeitos radioativos formados por substâncias cuja meia-vida pode alcançar milhares de anos, como um dos isótopos do plutônio, o 239, que tem meia-vida de 24 mil anos?

Imagine o que aconteceria, se a documentação sobre uma falha em um reator nuclear misteriosamente desaparecesse. Este foi o problema com que a Ontário Hydro [2] se deparou quando, no início de 1990, verificou que os registros digitais sobre a falha num componente não podiam ser recuperados. Muitos casos exemplares são repetidamente relatados pela literatura sobre a perda irreversível ou sobre o alto custo de recuperação de registros exclusivamente digitais de importância reconhecida, causados por práticas e políticas – ou ausência de políticas – equivocadas de gestão de repositórios digitais. São muitos os exemplos, mas entre os casos clássicos está o do censo americano de 1960, cujos dados foram armazenados em fitas que 16 anos depois só duas máquinas no mundo podiam ler, uma no Japão e outra guardada como peça de museu no Smithsonian Institution.[3] Uma operação de salvamento heróica e custosa recuperou a maior parte dos dados, mas não tudo (dizem que hoje se sabe mais sobre o censo americano de 1860 do que o de 1960) [4]. O Brasil também está bem representado: observações da Amazônia feitas por satélite na década de 1970 estão para sempre perdidas[14]. Essas informações foram perdidas por causa da obsolescência tecnológica: o *hardware* e/ou *software* usados para acessar ou ler os registros não estão mais disponíveis e a tecnologia atual não é compatível.

Deve ficar claro que o desafio de assegurar o acesso de longo prazo a informações armazenadas digitalmente coloca o profissional de informação diante de um problema de grande complexidade, posto que as questões envolvidas não se limitam ao domínio puramente tecnológico, o que já seria uma tarefa hercúlea. A gestão de informações digitais, devido principalmen-

te à sua transcendência temporal e à sua dependência aos compromissos de longo prazo, envolve ainda questões administrativas, legais, políticas, econômicas e financeiras e, sobretudo, questões referentes à representação das informações via metadados para acesso e gestão da preservação. Tudo isso, somado, coloca a preservação digital na classe dos problemas de natureza complexa. É exatamente disso que vamos tratar neste texto.

2. Bibliotecas digitais e a preservação digital

A preocupação sobre a preservação da informação eletrônica começou nos anos 70, dentro do ambiente da comunidade arquivística [14]. Durante esse tempo a arquivologia avançou bastante nessa área, tornando-se uma ciência mais tecnológica e mais sofisticada, incorporando conhecimentos novos da tecnologia da informação, principalmente os necessários para garantir a integridade e autenticidade dos objetos digitais sob a sua guarda – certificação, assinatura digital, criptografia, etc. Com isso caminhou também para superar o velho dilema que sempre a atormentou, que é o quase antagonismo entre a preservação e o acesso.

Somente nos últimos anos, com o potencial agravamento do problema, é que outros enfoques nesse sentido foram identificados, analisados e comparados, dentre os quais a ótica e perspectiva das bibliotecas digitais.

Mas a despeito dos pesados investimentos em tecnologia da informação, existe uma debilidade crítica na infra-estrutura dos sistemas de informação que os incapacitam de garantir preservação continuada e acessibilidade plena às informações geradas num contexto de rápido avanço tecnológico [5]. A maioria desses sistemas manipula informação digital segundo paradigmas do mundo analógico, gerando recursivamente enormes contradições e situações de risco para os objetos digitais sob suas custódias. A biblioteconomia – antes focada no padrão de sistemas de bases de dados referenciais e catalográficas – só tardiamente começou a incorporar metodologias de preservação digital aos seus processos, o que se deu com o surgimento das bibliotecas digitais e diante da enorme responsabilidade que isso acarreta para elas.

Não obstante esse distanciamento inicial das questões de preservação digital, as bibliotecas digitais hoje são geradoras, custodiantes e distribuidoras de materiais informacionais digitais de toda a natureza, e como

tal têm a responsabilidade inicial de assegurar a preservação de longo prazo e, conseqüentemente, a plena acessibilidade a esses materiais, posto que as condições para a criação de informação digital e para a atribuição de usabilidade a elas são essencialmente as mesmas requeridas para que elas perdurem. Há um consenso claro de que, à medida que os criadores/ disseminadores/ proprietários de informação digital aceitam a responsabilidade inicial de arquivarem seus objetos digitais, terão também de incorporar preservação digital nos seus processos [6].

Dentro desse contexto, as bibliotecas digitais se tornam cada vez mais um elo importante na perenização dos estoques de informação digital, as quais constituem testemunhos das atividades da organização ou sistema de organizações na qual essas bibliotecas estão inseridas. Um exemplo concreto disso são as redes de bibliotecas de teses e dissertações, cujos acervos digitais distribuídos em rede são o testemunho das atividades acadêmicas e de pesquisa de uma universidade, de um país ou de uma região [7,8]. Generalizando um grau a mais, pode-se verificar que cada vez mais instituições não-arquivísticas são obrigadas a assumir a custódia de estoques informacionais digitais de caráter arquivístico, fato já reconhecido pelos especialistas da área de arquivologia, tanto é que a norma ISO/OAIS [9] é dirigida também – e especialmente – para instituições não-arquivísticas, detentoras de responsabilidade como geradora, provedora ou custodiante, de informação digital.

As bibliotecas digitais sempre estiveram imersas em problemas de origem bem diversificada [7] – técnicos, gerenciais e econômicos – tais como banco de dados, armazenamento de massa, interoperabilidade, metadados, protocolos e padrões, direitos autorais e modelos de custo - e só agora, reconhecida a gravidade da questão e as responsabilidades de quem detém acervos digitais, voltam sua atenção para o problema de preservação digital e para o seu escopo também interdisciplinar e muitas vezes coincidente com os seus antigos problemas.

3. Preservação digital – complexidade e paradoxo

"A preservação digital envolve não somente a retenção do objeto informacional em si, mas também do seu significado. É necessário, portanto, que as técnicas de preservação sejam capazes de compreender e recriar

a forma original ou a função do objeto de forma que seja asseguradas sua autenticidade e acessibilidade" [10]. Além do mais, as estratégias de preservação devem ter uma amplitude tal que incorporem vários outros aspectos, incluindo custo-benefício, restrições legais e requisitos de acesso para o usuário-final. O próprio sentido conceitual de preservação, no contexto da informação digital, está imerso em um paradoxo: tradicionalmente preservar algo significa mantê-lo imutável e intacto; entretanto, no ambiente digital, preservar significa, na maioria dos casos, mudar, recriar, renovar: mudar formatos, renovar mídias, *hardware* e *software*. Por um lado, queremos manter a informação intacta, como ela foi criada; por outro, queremos acessá-la dinamicamente e com as mais avançadas ferramentas [5].

A longevidade da informação, finalmente, torna-se um problema importante e de interesse de muitas disciplinas, não somente da arquivologia e da biblioteconomia. Cada grupo com percepção, interesses e abordagens diferentes em relação à longevidade dos documentos digitais. Existem hoje numerosas iniciativas internacionais cuidando dessa questão. Entretanto, a variedade de soluções não tem implicado absolutamente um corpo de conhecimentos plenamente consolidados.

É de se esperar, portanto, que se tenham abordagens variadas de estratégias para a preservação digital. Elas vão desde a criação de museus tecnológicos, onde equipamentos e programas são depositados e mantidos; passando pelo desenvolvimento de programas emuladores de equipamentos antigos, que permitem – em tese – processar *softwares* obsoletos em qualquer computador do futuro; passando ainda pela migração, que se baseia no princípio de que os objetos digitais podem acompanhar a tecnologia por meio de atualização de mídias e formatos; chegando a uma abordagem que constitui uma autêntica regressão tecnológica: passar o material digital para meios mais estáveis – e analógicos –, como papel ou microfilme. A despeito da perda total dos atributos digitais, tais como hipertextualidade, som, movimento e distribuição em rede, esta é provavelmente a estratégia mais usada pelas organizações e pelas pessoas (ou você, leitor, nunca imprimiu uma foto digital achando que esta é melhor maneira de preservá-la?).

Nenhuma dessas estratégias tomada isoladamente é capaz de solucionar o espectro amplo de problemas decorrentes da fragilidade dos materiais digitais e de suas inúmeras dependências. Além do mais, a escolha

das estratégias é somente uma das facetas do problema de preservação do acesso aos documentos digitais. Temos que considerar ainda outras variáveis, como os aspectos tecnológicos decorrentes do uso intensivo de tecnologia da informação – *hardware*, *software*, armazenamento de massa, tecnologia Web, protocolos, normas etc.; os aspectos organizacionais, gerenciais, sociais, legais e jurídicos – gestão, planejamento, direitos autorais, autenticidade, legislação etc.; aspectos econômicos, cujo equacionamento ainda está nos primeiros passos – modelo de custo, modelo de negócio etc.; e principalmente os aspectos pertinentes à arquivologia e à ciência da informação. Essas diversas facetas interagem entre si pactuando compromissos de longo prazo, aumentando o número de variáveis do problema e colocando a preservação digital na classe dos problemas complexos. Concorre para isso a própria essência complexa dos objetos digitais informacionais, como será visto a seguir.

O que se espera da preservação digital é, em última análise, preservar o conteúdo intelectual de um objeto digital. Entretanto, a noção de conteúdo no mundo digital é estratificada e complexa. No mundo analógico, os documentos encerram em único suporte todos os seus atributos de forma monolítica: a presença física, o *layout*, o formato, o conteúdo e o suporte são elementos virtualmente inseparáveis, como é o caso de um livro convencional. Todo processamento acontece na cabeça do leitor, e quando o livro é preservado, todos os aspectos do livro são igualmente preservados. Em contraste, os objetos digitais são facilmente decompostos em elementos individuais, o que implica um esforço adicional – e significativo – para identificar quais elementos devem ser preservados para garantir que o objeto digital seja preservado como um todo. Bullock [11] identifica um conjunto de aspectos, enfatizando que as estratégias de preservação devem garantir que o maior número tecnicamente possível deles persista no tempo, quais sejam: limites do objeto digital; presença física – ou seja, a sua cadeia de *bits*; conteúdo – no seu nível mais baixo como um arquivo ASCII; apresentação; funcionalidades; autenticidade; localização e referência do objeto no tempo; proveniência – origem e cadeia de custódia; e contexto – relacionamentos e dependências de *hardware* e *software*.

Parece claro que o objetivo da preservação digital não é preservar os objetos digitais tão somente como artefatos físicos, independentemente da sua usabilidade, posto que para tal bastaria a guarda em ambiente

controlado das mídias que dão suporte a esses objetos. A preservação no universo digital tem pouca relação com a longevidade de discos óticos, fitas magnéticas e outros meios de armazenagem (bem guardado, o meu CD de música pode durar 100 anos, mas daqui a 25 anos haverá aparato para tocá-lo?). Hoje não há dúvida que o santo graal da preservação digital é a preservação do acesso contínuo ao conteúdo intelectual dos documentos digitais, e que a viabilidade das bibliotecas e arquivos digitais depende fortemente da expectativa de vida dos sistemas de acesso – um elo tão resistente quanto o seu mais frágil componente [12]. No mundo analógico – do papel e do microfilme – a preservação e o acesso são atividades relacionadas, porém distintas e muitas vezes antagônicas; no mundo digital a preservação e acesso são indissociáveis – a preservação digital se confunde com a própria preservação do acesso.

Conclui-se, portanto, que para manter os objetos digitais perenemente acessíveis para uso, requer-se algo mais do que preservar simplesmente o artefato físico, é necessário considerar também várias outras dimensões que o problema apresenta: 1) a preservação física, cujo foco está na preservação das mídias e na sua renovação quando se fizer necessário; 2) a preservação lógica, que tem como foco os formatos e a dependência de *hardware* e *software* que mantenham legíveis e interpretáveis a cadeia de *bits*; 3) preservação intelectual, que tem como foco o conteúdo intelectual e sua autenticidade e integridade; 4) é importante ainda considerar a preservação do aparato – na forma de metadados – necessário para localizar, recuperar e representar a informação digital; 5) assim como proceder ao monitoramento e à instrumentalização da comunidade-alvo, audiência para o qual a informação de forma privilegiada se dirige, no sentido de garantir que ele possa compreender plenamente a informação no momento do seu acesso.[13]

Os processos de preservação de informação digital variam enormemente em função dos diferentes tipos de objetos a serem preservados – texto, imagem, dados numéricos, vídeo, som, multimídia etc. Não importando a natureza do objeto digital, o objetivo fundamental é preservar a integridade do objeto, isto é, as suas características de objeto único e singular. Saber como se preserva, em termos operacionais objetos, codificados digitalmente, ou mesmo outros tipos de objetos, depende fortemente da capacidade de discriminar as características essenciais do que precisa ser preser-

vado. A integridade de objetos informacionais digitais não depende somente do objeto em si, mas está também vinculada aos diversos tipos de atributos que eles vão incorporando durante o curso de sua vida, bem como os interesses dos diversos atores envolvidos. Tudo isso somado é que finalmente lhes confere identidade distinta e singularidade [14].

Dois documentos têm importância fundamental na formalização e na conceitualização do problema da preservação dos objetos digitais e também no estabelecimento de um elenco de informações necessárias para instruir os processos de preservação. São eles: o relatório Preserving Digital Information [14] e a norma Open Archival information System (OAIS) [9].

Preserving Digital Information, relatório produzido pelo Task Force on Archiving of Digital Information em 1966, no âmbito da Commission on Preservation and Access (CPA) e do Research Libraries Group (RLG) estabelece que "no ambiente digital, as características que determinam a integridade da informação e merecem atenção especial para propósitos arquivísticos são as seguintes: conteúdo, permanência, referência, proveniência e contexto".

Por outro lado, a International Standard Organization – ISO – publicou em 2003 o modelo de referência para Open Archival information System – OAIS–[9], cujo desenvolvimento foi coordenado pelo Consultative Committe for Space Data System, de sigla CCSDS, e tinha como objetivo estabelecer padrões para dar apoio à preservação de longo prazo de informações digitais decorrentes de observações espaciais. Embora o desenvolvimento do modelo tenha se originado no âmbito da comunidade espacial, ele se dirige também a outras audiências. O documento define um modelo de referência de alto nível para um sistema aberto voltado para arquivamento de informações. No modelo o sistema de arquivamento é definido como uma organização de pessoas e sistemas que assumiu a responsabilidade de preservar a informação e torná-la disponível – em termos de acesso – para uma classe privilegiada de usuários, designada como comunidade-alvo. Entre outras coisas, o Modelo OAIS tem o objetivo de proporcionar uma arquitetura comum que pode ser usada para ajudar a compreender o desafio das organizações que tem a responsabilidade de arquivo, especialmente aquelas envolvidas com a informação digital e a sua preservação[15, 21].

Como parte dessa arquitetura, o Modelo OAIS identifica e distingue entre os diferentes tipos de informação – ou metadados – aqueles necessários

para dar suporte à preservação digital. Essas informações são divididas em quatro diferentes grupos fortemente baseados nos conceitos já estabelecidos pelo Relatório da CPA/RLG – informação de referência, informação de contexto, informação de proveniência e informação de permanência.

Conteúdo – Conforme já enfatizado, o objeto da preservação digital é o conteúdo, isto é, a substância intelectual contida nos objetos informacionais. No entanto, a noção de conteúdo é em si uma idéia complexa que opera em níveis distintos de abstração. Os objetos do reino digital precisam, para preservar a sua integridade, que se considere em que nível de abstração está definido o seu conteúdo informacional [14]. A natureza digital de uma informação é dada pela forma como ela é armazenada: seqüências de 0´s e 1´s, isto é, cadeias de *bits* (*bitstream*). Podemos, portanto, concluir que a forma mais básica de preservar a integridade de um objeto informacional é preservar a configuração de *bits* que o definem como objeto único. Mas esta estratégia é simplista demais e pouco útil, pois, "um arquivo, formado por cadeia de *bits*, não é um documento propriamente dito, ele simplesmente descreve um documento que se concretiza quando o arquivo é interpretado pelo *software* que o produziu. Sem esse *software* (ou um equivalente), o arquivo é um documento criptografado através do seu próprio código" [16]. Recuperada uma cadeia de *bits*, ela pode representar uma infinidade de coisas – de uma seqüência de inteiros a um arranjo de pontos de uma imagem, e deve, portanto, ter o seu significado interpretado. A maioria dos arquivos contém informações cujo significado só pode ser explicitado pelo *software* que os criou. Essa interpretação é fortemente dependente de formatos, códigos e estruturas – que não podem ser explicitamente representadas na cadeia de *bits* – e que são tratados por *software* e *hardware* específicos, que por sua vez estão sujeitos a um ciclo de obsolescência cada vez mais rápido. Estas dependências é que impõem as mais graves dificuldades na gestão da integridade da informação digital, pois pressupõem níveis sofisticados de controles[14, 16].

Portanto, para se compreender um documento é necessário conhecer o significado do seu conteúdo na linguagem para o qual ele foi construído. No caso de um documento digital, essa linguagem é um programa de computador. Dependendo da complexidade de um dado documento digital, pode-se extrair o seu conteúdo através de um *software* que não o original, com perdas de vários níveis. Um bom exemplo são os editores de texto que, via de regra, disponibilizam mecanismos para converter

documentos em formatos comuns de intercâmbio, porém o uso de tais mecanismos freqüentemente resulta em perdas e execução inadequadas em termos de estrutura, *layout* e funcionalidades.

O RLG [14] estabelece que, no mais alto nível de abstração, os arquivos definem os seus conteúdos de forma que transcendam os limites de *hardware* e *software* necessários para a leitura e interpretação dos *bits* de um objeto informacional e também para executá-lo segundo formatos e representações estruturais específicas; isto é, os arquivos devem definir o conteúdo em termos do conhecimento e idéias que o objeto contém. Portanto – ao menos por essa ótica, o desafio da preservação digital é fazer as conversões necessárias para a preservação do conteúdo intelectual usando algoritmos padronizados de intercâmbio e outras estratégias apropriadas, de forma que as idéias presentes no final sejam idênticas às contidas no objeto original.

Permanência (Fixity)

O Relatório CPA/RLG [14] declara que o processo de identificação e preservação de um objeto digital vai bem além das considerações sobre o seu conteúdo. Esse processo depende em grande escala também de outros fatores, dentre eles a forma como um conteúdo está fixado como um objeto discreto. "Se um objeto não está fixado e o seu conteúdo está sujeito a mudanças ou retificações não informadas, então sua integridade pode estar comprometida e o seu valor como registro cultural (e como valor de prova) pode estar também severamente reduzido". Na perspectiva do Modelo OAIS [15], o propósito da informação de permanência é assegurar que o conteúdo informacional armazenado em um repositório é autêntico e íntegro, posto que a mutabilidade dos conteúdos em forma digital torna esse problema um dos mais críticos no contexto da preservação digital. No entanto, há que se destacar que essa questão é minimizada pelas diversas técnicas de autenticação e criptografia hoje disponíveis – marca d´água, assinatura digital, certificação digital, algoritmos verificadores etc., mas é preciso ainda desenvolver consideravelmente padrões, políticas e metodologias para a aplicação adequada dessas técnicas na geração de objetos digitais convenientemente fixados. Em termos de metadados, a informação de permanência documenta os mecanismos de autenticação usados para assegurar que o conteúdo

informacional de um objeto digital não sofreu alterações não documentadas; ela se refere a qualquer informação que documenta mecanismos de autenticação em uso em um dado repositório.

Referência

Para que um objeto mantenha sua integridade e as características que o identificam como um objeto único e singular, é necessário também que se possa localizá-lo de forma definitiva e confiável entre outros objetos ao longo do tempo. Isto nos indica que é necessário que o conteúdo informacional seja, de algum modo, identificado e descrito[14]. Portanto, um meio consistente de referência a um objeto informacional constitui um aspecto importante no processo de preservação digital. Parte deste problema geral - identificar e corretamente referenciar objetos digitais – consiste no problema específico de resolução de nomes e localização, que é objeto de iniciativas importantes, principalmente no âmbito World Wide Web Cosortium – W3C (http://www.w3c.org) – entidade responsável pelo ordenamento da Internet.

O método mais usado para indicar a localização de um objeto digital no mundo Web é a Uniform Resource Locator (URL). Ela se refere a um lugar onde o objeto reside, mais especificamente, em que computador servidor ele está armazenado e disponível para acesso. A debilidade da URL é que ela varia com muita freqüência, especialmente quando o objeto migra de uma máquina para outra. Outros métodos mais robustos têm sido testados pela W3C e por outras organizações: Uniform Resource Name (URN)[17], que estabelece um nome único e permanente para cada objeto; Persistente URL (PURL)[18] e Digital Object Identifier (DOI)[19].

Informação de Referência, nos termos do OAIS [9], descreve sistemas de identificação e mecanismos que permitem assinalar identificadores, usados para identificar univocamente o conteúdo informacional, tanto interna como externamente, em relação ao sistema de arquivos onde ele se encontra armazenado. Dessa forma, a Informação de Referência também permite que sistemas externos façam referência a este conteúdo informacional particular, o que sugere duas principais funções para a informação de referência: em primeiro lugar ela deve identificar o conteúdo informacional localmente – ou seja, dentro do sistema de arquivo que ele reside; em seguida ela deve identificar o conteúdo informacional globalmente, isto é, para um sistema

externo ao sistema de arquivo onde está armazenado o conteúdo informacional. Um bom exemplo para esta função dupla da informação de referência é a catalogação de um livro que é identificado dentro do contexto de sua coleção, via número de chamada e no contexto global – do universo de livros – por uma agência externa via seu ISBN, ou pelo Digital Object Identifier (DOI). Um nível intermediário de identificação pode ser observado no contexto da catalogação cooperativa, por exemplo, por um número OCLC. A informação de referência pode também ser usada para armazenar informações descritivas que podem ser usadas por usuários para descobrimento de recursos. Nesse sentido, muitas iniciativas em arquivos digitais estenderam a definição de informação de referência para incluir também dados descritivos. O OAIS observa que no contexto das bibliotecas digitais, informações de referência podem também incluir descrição bibliográfica.

Proveniência

O Relatório da CPA/RLG [14] considera que a proveniência tem se tornado um dos conceitos organizacionais centrais da ciência arquivística moderna. O seu texto indica que uma suposição subjacente ao princípio da proveniência é que a integridade de um objeto informacional está parcialmente incorporada na sua trajetória, desde sua origem, e que para preservar a integridade de um objeto informacional, os arquivos digitais devem necessariamente preservar o registro de sua origem e a sua cadeia de custódia. Ainda segundo os resultados da Força Tarefa da CPA/RLG, o investimento em estabelecer a proveniência dos objetos informacionais é importante para preservar a integridade deles de duas maneiras distintas. Em primeiro lugar: traçar a cadeia de custódia a partir do ponto de criação de um dado objeto ajuda no estabelecimento da suposição que ele é autêntico, e que ele é o que se supõe que ele seja, e que seu conteúdo não foi manipulado, alterado ou falsificado [20]. Em segundo lugar: estabelecer a proveniência através da cadeia de custódia significa documentar, pelo menos em parte, o uso específico do objeto pelos seus custodiantes ao longo do tempo, ou seja, criar um registro de uso para o objeto.

Nos termos da taxonomia do OAIS [9], informação de procedência refere-se geralmente à informação que documenta a cronologia do conteúdo informacional, incluindo sua origem ou fonte, e quaisquer mudanças do seu objeto ou do seu conteúdo sofrida ao longo do tempo desde sua

origem – criação, transferência de propriedade, migração de formatos; inclui também informações sobre a cadeia de custódia, isto é, quem tem tido sua custódia desde que ele foi originado. A informação de procedência pode registrar ainda informações sobre propriedade intelectual, permissão de acesso, depósito legal, dentro do escopo da gestão de acesso do conteúdo informacional.

Contexto

O atributo contexto de um objeto informacional digital está relacionado à forma como ele interage com os outros elementos do amplo ambiente digital. O modelo de referência OAIS [9] define informação de contexto como sendo "a informação que documenta o relacionamento do conteúdo informacional com o seu meio ambiente. Isto inclui a motivação de sua criação e o seu relacionamento com os outros objetos". De acordo com o Relatório CPA/RLG, o contexto de uma informação digital inclui uma dimensão técnica, uma dimensão de *linkage* com outros objetos, uma dimensão de comunicação e uma ampla dimensão social.

O contexto técnico especifica as dependências dos objetos digitais relativas ao *hardware* e ao *software*, que, devido a sua natureza intrínseca, requerem o uso de computadores e programas tanto para criá-los assim como para usá-los. O nível de dependência pode variar enormemente, desde um objeto que depende de processadores e sistemas operacionais específicos, até os documentos codificados em linguagens de marcação, tais como HTML, SGML e XML, que não têm dependências específicas. Isto coloca mais um desafio para as instituições cuja missão é preservar a integridade dos objetos informacionais digitais – os arquivos e bibliotecas digitais –, que é representar o mais fielmente possível o contexto dos objetos em relação à dependência de *hardware* e *software*; e superar, por meio de mecanismos apropriados de migração – ou de outras estratégias de preservação –, os problemas que ameaçam o uso futuro dos materiais digitais.

A dimensão *linkage* do contexto dos objetos informacionais digitais constitui um problema de extrema complexidade e ainda sem solução satisfatória, posto que impõe para um objeto digital residente em um ambiente de rede, e cuja integridade depende muito mais da suas interrelações com outros recursos do que no objeto individual, que os meca-

nismos de preservação contemplem não somente o objeto isoladamente, mas também seus *links*. Uma *homepage*, por exemplo, para ter a sua integridade preservada, deve ser preservada em conjunto com seus *links* e com outros objetos.

O contexto pode também incluir informações relativas ao modo de distribuição de um dado objeto digital – se ele, por exemplo, reside numa rede ou está armazenado num CD-ROM –, constituindo a dimensão comunicação. Como objetos digitais são, com freqüência crescente, gerados e mantidos puramente em ambiente de rede, as características da rede, tais como segurança e largura de banda, tornam-se importante na sua caracterização. Dessa forma, a integridade de um dado objeto digital distribuído via rede deve levar em conta as características da rede que suportam a sua existência.

O Relatório da RLG também toca numa dimensão social do contexto, que teria contribuições à integridade de objetos informacionais digitais. Esta dimensão poderia incluir, por exemplo, as políticas específicas e decisões de implementação que definem os vários atributos de uma rede: largura de banda, segurança, restrições de uso etc.; e o contexto social e organizacional onde os registros são gerados. São fatores que podem caracterizar o interesse das pessoas envolvidas – os *stakeholders* – nos objetos informacionais.

4. As estratégias de preservação digital

Até o presente momento – e provavelmente até um futuro indefinido – não teremos disponível uma estratégia única capaz de dar conta de todo o espectro de problemas relacionados à preservação digital. O que se apresenta são soluções específicas para casos específicos. Como veremos a seguir, são muitas as metodologias de preservação digital: algumas apresentam soluções parciais; outras são simplesmente ausência de solução que partem principalmente dos que confiam demasiadamente que as tecnologias do futuro e as técnicas da arqueologia digital – trabalho sofisticado de recuperação de objetos digitais danificados ou inacessíveis pela obsolescência tecnológica – serão capazes de recuperar o que pode já estar para sempre perdido. Dentre as soluções potencialmente completas, algumas são de uso corrente, mesmo que em pequena escala,

outras são experimentais e estão ainda nas bancadas dos laboratórios ou ainda em patamares bastante teóricos.

De forma geral, a comunidade envolvida no problema de preservação digital foi capaz de desenvolver métodos efetivos para a preservação do conteúdo informacional de materiais digitais, cujos formatos e estruturas são bem conhecidos e mais simples, mas falha ou obtém resultados pífios para materiais mais complexos e dinâmicos, ou para os que constituem formas novas ou emergentes de documentos e que expressam a inovação no uso da tecnologia digital.

Por mais surpreendente que seja, uma das estratégias mais comumente usadas no afã de se conservar o conteúdo intelectual de documentos digitais é fixá-lo em suportes analógicos, mesmo tendo-se em conta as perdas óbvias dos seus atributos digitais, tais como apresentação, funcionalidades, distribuição em rede, hipertextualidade e hipermídia. A forma mais elementar de se fazer isto é imprimir os documentos digitais em papel, que é ainda mais longevo que os equivalentes digitais, ou melhor ainda, fixá-los em microfilme, onde a estabilidade e valor de prova poderão estar garantidos por séculos. Papel e microfilme são mais estáveis que a maioria das mídias digitais e não necessitam de *software* e *hardware* especiais para recuperar os seus conteúdos.

Explorando esse mesmo caminho, novos produtos comerciais são lançados tendo por base gravações analógicas. Este é o caso do "HD Rosetta" [22] que promete aos seus clientes mil anos de duração para as páginas gravadas em suas placas de metal e recuperadas com auxílio de microscópios óticos ou eletrônicos. Estas soluções constituem verdadeiras regressões tecnológicas, à medida que se valem de suportes tecnologicamente ultrapassados para garantir um futuro cujo limite não podemos avaliar, – a preservação de artefatos digitais.

Ainda seguindo a estratégia de regressão tecnológica, se pode converter os documentos para formas digitais mais simplificadas, tais como texto puro – formato ASCII; dessa forma, minimizando a exigência de *softwares* de recuperação sofisticados, podem atravessar sucessivas gerações de tecnologia. Pode-se ainda transformá-los em imagens em formato digital – formatos PDF, TIFF, JPEG etc. – a partir de documentos digitais originais. Para ambos os casos persistem, todavia, os problemas de preservação digital dos objetos resultante das conversões, e somam-se ainda as perdas

severas dos atributos inerentes aos documentos digitais, principalmente no caso dos documentos com características multimídias ou formatos mais complexos. Entretanto, estas são estratégias de baixo custo e especialmente viáveis pela sua facilidade e universalidade, para o caso em que reter o conteúdo é o mais importante, e que outras funcionalidades presentes – exibição, indexação, características computacionais etc. – não são críticas. O fato real é que enquanto perdurar a carência de estratégias de preservação mais robustas, consolidadas e de custo-benefício favorável, a impressão em papel, microfilme ou em outros dispositivos analógicos, bem como a conversão para formatos mais simples permanecerão como as únicas estratégias viáveis para muitas pessoas e organizações [14, 23, 24]

Considerando que uma cadeia de *bits* faz parte da camada mais elementar dotada de um significado intencional de um objeto digital, pode se supor, a princípio, que o requisito básico para a preservação digital é assegurar que esta cadeia de *bits* estará para sempre armazenada de maneira íntegra num meio estável. Pode se supor ainda que se este meio se deteriora ou se torna obsoleto antes que a informação tenha sido transferida para outra mídia, a informação estará perdida e portanto, a preservação digital está circunscrita à cópia da informação digital para uma nova mídia, antes que a informação armazenada na mídia atual se torne obsoleta e não possa ser mais acessada.

A realidade, porém, mostra que a preservação digital é algo mais complexo do que a simples cópia para uma mídia mais atualizada. Isto é necessário, mas raramente suficiente como estratégia única de preservação digital e como garantia de que a informação possa ser recuperada e processada com a tecnologia disponível no futuro. O acesso e interpretação de uma cadeia de *bits* requerem estruturas gerenciais e tecnológicas sofisticadas e metainformações que orientem a extração do seu significado e a compreensão plena, a qualquer momento, pela comunidade-alvo para a qual o objeto foi originalmente produzido.

Padrões

Um fator fundamental de sucesso para o aumento da longevidade dos objetos digitais, não importando a estratégia de preservação digital adotada, está relacionado com a adoção de padrões, especialmente os padrões abertos – que são padrões cuja documentação está publicada e pronta-

mente disponível, e que tem o desenvolvimento baseado em processos consensuais envolvendo as várias partes interessadas. Numa visão otimista, os padrões abertos permitem que os documentos digitais sejam representados em formatos mais duradouros e estáveis, dessa forma reduzindo a velocidade do ciclo de obsolescência dos objetos digitais. São os padrões que formam as bases de funcionamento das bibliotecas digitais, e que tornam possíveis muitas de suas funcionalidades essenciais, como por exemplo, a interoperabilidade entre sistemas e intercâmbio automático de informações[28]. A aplicação de padrões na preservação digital – na codificação, nos formatos e nos esquemas de representação – torna os processos de preservação digital mais fáceis, menos freqüentes e mais baratos, à medida que reduzem a grande variedade de processos de preservação customizados, que são decorrentes da multiplicidade de formatos em que se traduzem os objetos digitais não padronizados. Idealmente a padronização deve preceder a própria criação do objeto da preservação[14, 23, 24].

Esta abordagem se baseia no pressuposto de que produtos da indústria de tecnologia da informação – *hardware*, *software* e mídias – que são aderentes com padrões largamente utilizados, estão mais difundidos no mercado e são, portanto, menos voláteis [23].

É necessário enfatizar que os padrões de uso corrente, via de regra, não foram planejados especificamente para a preservação digital e talvez por isso não funcionem tão bem, mas facilitam a comunicação e a ação mútua entre os vários domínios – conceituais, tecnológicos, representacionais etc. – que concorrem para a preservação digital e são considerados essenciais para tal. Numa primeira análise, podemos diferenciar as seguintes classes de padrões de interesse [25]:

- padrões que servem como modelos de referência (arquitetura) descrevendo funcionalidades, procedimentos, fluxos de informação, conceitos concernentes a bibliotecas e arquivos digitais. São exemplos dessa categoria a ISO/DIS 15489, DoD 501.2-STD e o OAIS, que serão especialmente discutidos mais adiante;
- padrões voltados para preservar o formato do objeto digital, a apresentação e funcionalidades e para facilitar o intercâmbio de documentos. Por exemplo: XML e PDF;
- metadados padronizados para preservar o acesso ao documento e

descrever o conteúdo, o contexto tecnológico, a proveniência e o significado, permitindo a recuperação e a interpretação futura dos documentos digitais. O conjunto de metadados preconizados pelo OAIS para preservação, Dublin Core, MARC e ISAD(G) são alguns exemplos.

Devemos, entretanto, dedicar atenção a algumas limitações inerentes à utilização de padrões como estratégia de preservação digital: em primeiro lugar os padrões e protocolos, mesmos os abertos, se tornam obsoletos e sofrem mudanças decorrentes do desenvolvimento vertiginoso das áreas de tecnologia da informação e das mudanças de paradigmas – por exemplo, em banco de dados: modelo de dados hierárquico para modelo relacional e em seguida para o modelo orientado a objetos –, o que indica que eventualmente haverá necessidade de converter materiais digitais em formatos obsoletos, porém, padronizados, para novos formatos. Nos casos de rápido desenvolvimento, parece ser irreal esperar que os padrões tenham a longevidade desejada; em segundo lugar – também em decorrência dos avanços tecnológicos – há muitos setores nos quais padrões técnicos inexistem, mesmo os proprietários, ou são conflitantes, o que acontece especialmente no caso de novos tipos de mídia e novas formas de representação. Enquanto os padrões são bem definidos para texto (exemplo: ASCII), imagens (exemplo: TIFF e JPEG) e documentos codificados (exemplos: XML, HTML), não se observou o surgimento de padrões consolidados para outros tipos de informação, como para bases de dados [11]; finalmente, na ausência de padrões abertos para alguns aspectos de um dado objeto digital, padrões proprietários se tornam padrões *de fato*, isto pode acontecer mesmo quando padrões abertos existem, mas são tecnicamente inferiores aos padrões proprietários correspondentes [26].

Apesar dessas limitações, o uso de padrões como estratégica básica de preservação digital não deve jamais ser descartado, posto que em algumas áreas eles cumprem um papel importante. Este é precisamente o caso das bibliotecas e arquivos digitais cujas coleções estão centradas em materiais textuais e iconográficos que não são, via de regra, dinamicamente interativos, tais como relatórios, periódicos e documentos eletrônicos e coleções de imagens, que por sua natureza podem ser beneficiados com adoção de padrões já consolidados, tais como a *Extensible Markup Language* – XML –, e o Tagged Image File Format – TIFF.

Existe, portanto, um consenso claro entre os especialistas da área de que a preservação digital pode ser realmente facilitada com a adoção de alguns procedimentos que incluem: 1) a definição de um conjunto limitado e gerenciável de padrões, preferencialmente abertos e/ou de ampla aceitação e de uso corrente; 2) a aplicação desses padrões na criação de novos objetos digitais, ou na conversão de documentos analógicos para formatos digitais; 3) o acompanhamento da obsolescência dos padrões desse conjunto e o monitoramento do surgimento de novos padrões; 4) a migração para os novos padrões tão logo eles estejam consolidados [11].

Portanto, o uso de padrões será de grande utilidade se houver consenso entre os vários atores envolvidos na questão da preservação digital sobre que conjuntos de padrões são potencialmente eficazes para determinados tipos de materiais digitais; se a indústria de T.I. tornar prontamente disponíveis produtos e ferramentas de *software* que estejam em conformidade ou dêem sustentação a esses conjuntos de padrões; e ainda se os gestores de arquivos e bibliotecas digitais puderem exigir aderência aos padrões como condição para inclusão de documentos nesses repositórios.

As abordagens de preservação digital

Lee [10] classifica as metodologias de preservação digital existentes em duas grandes abordagens. A primeira delas é mais conservadora e procura adotar como estratégia a restauração plena do ambiente tecnológico original para decodificar os objetos digitais no futuro. Isto pode ser realizado de duas formas: preservando-se réplicas reais de plataformas de *hardware* e *software* para uso no futuro – chamada de estratégia de preservação da tecnologia; ou recriando-se virtualmente nos novos computadores, por meio de programas emuladores, as plataformas obsoletas. Esta estratégia é conhecida como emulação da tecnologia. Em poucas palavras, esta abordagem está baseada na criação de museus tecnológicos, ainda que eles sejam virtuais e os seus acervos sejam criados por programas.

A outra vertente identificada por Lee procura superar a obsolescência dos formatos dos arquivos por meio de duas estratégias: a primeira delas se baseia na conversão de formatos e na renovação e transferência de mídias, e é conhecida como estratégia de migração; a outra estratégia é chamada de encapsulamento, onde tudo que é necessário para o acesso

ao objeto digital é agrupado fisicamente ou logicamente, sendo então preservado, incluindo informações que são expressas por metadados.

Migração

A estratégia de migração envolve basicamente um conjunto de atividades – que devem ser repetidas periodicamente – que consiste em copiar, converter ou transferir a informação digital do patamar tecnológico que a sustenta – mídias, *software*, formatos e *hardware* – para um outro mais atualizado e de uso corrente. Analisando com um grau a mais de profundidade, veremos que migração envolve transferir a informação digital de uma mídia que está se tornando obsoleta ou fisicamente deteriorada, ou instável, para um suporte mais novo ou tecnologicamente mais atualizado, por exemplo: de um CD-ROM para outro CD-ROM mais novo, de disquete para um DVD; de um formato ultrapassado para um formato mais atual ou para um formato padronizado, por exemplo: de MSWord para XML ou ASCII; de uma plataforma computacional em vias de descontinuidade para outra mais moderna, por exemplo: de VAX para UNIX [10,14].

A migração é a estratégia de preservação digital mais utilizada em maior escala; é também a que possui um conjunto de procedimentos mais bem organizados e consolidados dentre as estratégias correntemente praticadas. Talvez seja importante lembrar que migração já vem sendo praticada há muito tempo, antes mesmo de a questão da preservação digital se tornar um problema crítico para a sociedade. Os arquivos de dados – responsáveis pela guarda dos dados computacionais de uma organização, por exemplo – possuem uma longa história de uso com sucesso das técnicas de migração [11]. Este sucesso ocorre especialmente porque estes repositórios lidam com informações relativamente homogêneas e depositadas segundo procedimentos e formatos, sempre que possível, mais simples e/ou padronizados.

O objetivo primordial da migração é preservar a integridade de objetos digitais enquanto mantém – prioritariamente – a capacidade do usuário de recuperá-los, exibi-los e utilizá-los em face da constante mudança tecnológica. O seu foco está centrado no conteúdo informacional imbricado em um objeto digital, em detrimento da tecnologia que o envolve; e, não menos importante, na possibilidade de assegurar a acessibilidade futura a esse conteúdo, usando a tecnologia disponível à época do acesso. Com

essas características, a migração preserva potencialmente o conteúdo de um objeto digital; entretanto, ela pode não ser capaz de preservar algumas das funcionalidades mais complexas, como por exemplo, as propriedades multimídias e o contexto – *links* e outros relacionamentos – do objeto digital. E mais: migrações sucessivas podem eventualmente levar a perdas severas e inaceitáveis; a idéia básica, portanto, é minimizar essas perdas e reter o conteúdo numa forma mais usável possível [11, 14, 24].

As estratégias de migração e os seus custos associados variam consideravelmente, posto que são fortemente dependentes da sustentação tecnológica do material digital a ser preservado, da diversidade de formatos, do grau de complexidade dos atributos de exibição e das facilidades de acesso e recuperação que se deseja mantida através do tempo. O que se observa é que os procedimentos de migração aplicados a informações digitais assentada em arquivos de dados mais simples estão bem estabelecidos e em funcionamento operacional, mas as soluções para objetos digitais mais complexos, só agora estão nas primeiras etapas do seu encaminhamento pela comunidade envolvida com os problemas da preservação digital. Essas soluções exigem ainda um grande esforço de pesquisa e interdisciplinaridade; de estudos de viabilidade técnica, econômica, legal e gerencial; de análises comparativas e de avaliação sobre as melhores práticas [14].

Os pesquisadores da área vem estabelecendo taxonomias para as diferentes abordagens da migração, baseadas principalmente na medida da intervenção da migração sobre o objeto digital, mais precisamente, no grau de transformação que sofre a cadeia de *bits* original do objeto digital submetida a um determinado processo de migração. Alguns autores consideram também para fins de classificação os efeitos sobre as funcionalidades e *look and feel* do objeto original, e ainda fatores tais como: risco associado, esforço humano e complexidade técnica envolvida [10]. A norma OAIS [9] identifica quatro tipos de migração, que têm ainda como princípio de divisão o grau de interferência da migração sobre a integridade da cadeia de *bits*: rejuvenescimento (*refreshing*); replicação, reempacotamento e transformação. No rejuvenescimento, a mídia em vias de deterioração é simplesmente substituída por outra, assegurando-se uma cópia perfeita da cadeia de *bits* original; na replicação a cadeia de *bits* é mantida, mas ocorrem mudanças nos mecanismos de mapeamento das informações; enquanto no reempacotamento a interferência ocorre ao ní-

vel das metainformações de empacotamento do objeto, conforme definida pelo modelo OAIS; finalmente na transformação há uma real interferência na cadeia de *bits* do objeto digital.

Assim como o uso de um pequeno número de padrões evita que se tenha que fazer um número excessivo de migrações customizadas, além de reduzir o ciclo temporal de operações de migração pelo alongamento da longevidade dos objetos digitais, a compatibilidade retrospectiva dos *softwares* seria outro fator de impacto positivo para a comunidade que se interessa pela preservação digital. A compatibilidade retrospectiva permite que *softwares* atuais leiam dados de sistemas mais antigos sem necessidade de reformatações substanciais e sem perdas de funcionalidades. O que se propõe é que compatibilidade retrospectiva faça parte do elenco de facilidades padrão dos pacotes de *software*.

Embora as estratégias de migração venham rapidamente se tornando mais efetivas e mais consolidadas, fortalecidas pela experiência prática adquirida pelas comunidades envolvidas com o problema, e pelo aprendizado de como selecionar métodos mais efetivos e apropriados, a migração permanece em grande escala experimental e fornece um terreno fértil para pesquisa e esforço de desenvolvimento [14].Contudo, as fragilidades e incertezas da migração ficam mais evidentes à medida que crescem o volume e a diversidade dos estoques digitais que necessitam de gestão para a preservação de longo prazo. Quando alguém assume a responsabilidade de preservar um objeto digital, pode ser difícil predizer quando a migração será necessária, o que será requerido, quantas reformatações terão que ser feitas, o quanto de informação e funcionalidade será perdido, quanto a migração irá custar, ou – o que é pior – se ela é possível [23]. O que se tem certeza é que as operações de migração, a cada ciclo, terão de ser repetidas para cada um dos documentos, mesmo que esses documentos sejam raramente acessados e que cada formato, cada aplicação, cada tipo de documento vá exigir uma operação específica, o que torna virtualmente impossível a adoção de métodos automatizados.

Mas as críticas à migração como estratégia de preservação digital não se esgotam no campo da gestão e custo. A migração envolve quase sempre interferência na cadeia de *bits* – exceto quando envolve a simples cópia da cadeia de *bits* original para uma nova mídia –, cujas implicações podem significar perdas severas de informação, de funcionalidades, in-

trodução de erros e modificação na forma como a informação é exibida para o usuário. Migração é a cópia da cópia, da cópia que mantém o acesso, mas não preserva os originais. Por fim, as técnicas de migração podem não ser capazes de converter objetos complexos como as multimídias e documentos criados a partir de novos paradigmas técnicos, culturais e artísticos [24].

Preservação da tecnologia

Esta estratégia pressupõe que um museu de equipamentos e programas – plataforma de *hardwares* e periféricos, sistemas operacionais, *drivers* e o programa de aplicação original – podem ser preservados com a finalidade de replicar no futuro a configuração necessária para recuperar um objeto digital no seu ambiente original [10].

A vantagem imediata desse método é que o objeto digital mantém todas as suas propriedades, aparência e comportamento originais, visto que será processado no seu ambiente nativo. Há um certo grau de consenso que esta pode ser a melhor estratégia para certos objetos num horizonte temporal curto, quando a migração não é possível. Porém, num prazo mais longo ela se torna absolutamente proibitiva em termos de suporte técnico, espaço e custo [11, 26].

É fácil compreender as dificuldades de manter e gerenciar museus tecnológicos, se pensarmos na velocidade em que gerações de *hardware* e *software* se multiplicam e no custo – quando possível – de manutenção e assistência técnica; no problema da documentação e na deterioração dos equipamentos, pois é improvável que as máquinas funcionem indefinidamente, considerando tão somente que os próprios chips estão submetidos a decaimentos inerentes à sua natureza física.

Além do mais, é necessário compreender que os objetos digitais antigos e os *softwares* aplicativos que os suportam raramente poderão sobreviver nas suas mídias originais, e terão de ser copiados para mídias atualizadas, o que os impede de serem lidos em seus equipamentos originais. Isto implica que interfaces entre cada máquina antiga e novas gerações de mídias deverão ser desenvolvidas [26].

Por último, os documentos recuperados por essa estratégia têm o acesso restrito aos museus, perdendo os seus atributos de distribuição por rede.

Emulação

Esta estratégia está bem próxima à filosofia da preservação tecnológica, tendo em vista que envolve preservar programas aplicativos originais, os objetos digitais originais e todas as suas funcionalidades. A emulação parte do pressuposto que é tecnicamente viável substituir as plataformas de *hardware* obsoletas necessárias para rodar no futuro os aplicativos originais por máquinas virtuais, por meio de programas emuladores. Em vez de museus de equipamentos reais, teríamos museus virtuais, constituídos de programas emuladores de *hardwares* obsoletos.

Mas antes de prosseguir, é necessário compreender que um emulador é algo que desempenha a função de outro, e que no contexto da tecnologia de informação um emulador é um programa que permite que um equipamento desempenhe o papel de outro – e diferente – equipamento. Isto possibilita que um programa emulador de um computador obsoleto possa rodar num computador do futuro, permitindo que neste computador se processem programas que originalmente rodavam no computador obsoleto [26].

A emulação tem estado sob muitas pesquisas e controvérsias. É considerada por muitos pesquisadores como uma opção viável para superar as fragilidades da migração, ao mesmo tempo em que sofre severas críticas de outros estudiosos da área. Rothenberg, que empunha a bandeira da emulação, propõe um modelo que preserva intacto o documento digital original – ou seja, a sua cadeia de *bits* –, o *software* de aplicação original usado para criar e/ou recuperar o documento e o sistema operacional apropriado e outros arquivos necessários para restabelecer o ambiente de processamento do *software* de aplicação. Tudo isto deve ser encapsulado em um "envelope", que deve conter ainda especificações que permitam, numa data indefinida no futuro, a construção de um emulador, e documentação – que deve permanecer legível para humanos – com instruções para acesso ao envelope e interpretação do documento objeto da preservação, incluindo metadados de proveniência, de contexto e administrativos [26].

As críticas colocadas se concentram na complexidade técnica de desenvolver emuladores confiáveis e o alto risco que isso acarretaria para os objetos digitais dependentes dessa solução. A abordagem da emulação requer a disponibilidade da especificação detalhada do equipamento obsoleto, que por várias razões nem sempre é possível. Espera-se que o

desenvolvimento de formatos padronizados para a especificação de *hardware* facilite no futuro essa tarefa [14].

Apesar das controvérsias, percebe-se um consenso em torno de algumas possibilidades importantes da emulação, dentre elas a possibilidade que oferece de manter – pelo menos a curto e médio prazo – a presença física do objeto digital e suas funcionalidades originais, que pode ser crítica para alguns objetos e alguns usuários. A estratégia também é potencialmente útil para o caso em que a migração não é possível, seja porque se desconhece o formato do objeto, seja pela sua complexidade, seja porque o recurso é fortemente dependente de *hardware* e *software* particulares ou ainda porque não se pode convertê-lo para formatos independentes de *software* [10, 14].

Pode-se argumentar também que emulação é mais "econômica" do que a migração – apesar de essa diferença não poder ser minimamente quantificada – porque um programa emulador de uma configuração de *hardware* particular pode ser usado para acessar toda a informação digital que requer tal configuração. Entretanto, necessita-se ainda muita pesquisa e comparações para que a emulação possa finalmente sair da bancada dos laboratórios.

Encapsulamento

A estratégia de encapsulamento envolve criar "containers" ou "embrulhos" – estruturas físicas ou lógicas – onde todos os elementos necessários para a interpretação do objeto digital estão reunidos, incluindo o próprio objeto e os metadados correspondentes. Os metadados estabelecidos pelo Modelo de Referência OAIS também são usados para fornecer as informações de suporte que devem estar presentes em um encapsulamento: informação de representação, usadas para interpretar corretamente a seqüência de *bits*, contexto, proveniência, referência e permanência.

O Universal Preservation Format (UPF) [27] é um mecanismo de armazenamento de arquivo de dados autodescrito, que utiliza a estrutura de *wrappers* para encerrar o objeto digital e metadados com o propósito de protegê-los contra a obsolescência tecnológica.

5. À guisa de conclusão

Reconhecido o caráter efêmero dos objetos digitais, as bibliotecas e arquivos digitais devem ser pensados a partir de conceitos, procedimen-

tos e políticas que tornem tão longevos os seus estoques informacionais quanto eles são demandados pela sociedade de hoje, tendo como perspectiva o amanhã e o depois.

Portanto, como conclusão talvez seja mais útil estabelecer um conjunto de recomendações, extraídas do texto, que sintetizem as ações necessárias para se construir um corpo de conhecimento e um elenco de práticas voltadas para a preservação de documentos digitais.

Padrões. Um dos desafios importantes é criar diretrizes, políticas e padrões que o governo, empresas, a ciência e tecnologia possam usar hoje para assegurar a longevidade dos registros digitais. Isto inclui promover o uso de padrões e protocolos abertos, estáveis e de uso amplo; adotar padrões na criação, armazenamento e transmissão de documentos digitais; monitorar o surgimento de novos padrões e migrar quando necessário. O uso de padrões torna mais fácil e menos dispendiosa a aplicação de estratégias de preservação. Arquivos e Bibliotecas digitais devem manter-se à frente do desenvolvimento de padrões e garantir que sua própria infra-estrutura esteja em conformação com os padrões largamente adotados.

Gestão documental. Aplicar procedimentos e estratégias de gestão documental quando da criação, tratamento, transmissão e preservação de documentos em formatos digitais; desenvolver planejamento de longo prazo; definir critérios para seleção do patrimônio digital, pois nem tudo poderá ser salvo; desenvolver modelos de custo e de minimização dos riscos das conversões; promover o uso e o estabelecimento de procedimentos de melhores práticas. Para ajudar os bibliotecários e arquivistas do futuro, nós precisamos de métodos para organizar a informação digital hoje.

Tecnologia. Aplicar as técnicas apropriadas de preservação digital respeitando as especificidades de cada problema, pois não há – até a presente data – solução genérica e nem solução que funcione em grande escala; pesquisar e desenvolver soluções abertas em cooperação com a indústria de tecnologia da informação; e, sobretudo, monitorar as inovações tecnológicas potencialmente úteis e validá-las quando apropriado.

Segurança digital. Apoiar o uso de instrumentos oriundos da tecnologia da informação e de aceitação consensual que garantam a integridade, a confiabilidade e a autenticidade dos documentos digitais, e

que os proteja contra acidentes e intervenções não autorizadas, tais como, certificação digital e criptografia.

Metadados. Incentivar o uso de estruturas padronizadas de metadados orientadas para a gestão da preservação digital, para a acessibilidade dos documentos digitais e para a interoperabilidade dos sistemas.

Agenda de pesquisa. Desenvolver uma agenda nacional de pesquisa orientada para os problemas da preservação do acesso e da longevidade digital, alinhada com as principais iniciativas internacionais. Além das considerações naturais provenientes da arquivísticas, da biblioteconomia e da tecnologia da informação, é necessário incluir aspectos legais, sociais e éticos.

Legislação. Dispor de um corpo de leis que garanta a proteção do patrimônio digital e o seu reconhecimento pleno como valor de prova. Isto envolve os problemas de depósito legal e as restrições impostas pelas questões de *copyright* e propriedade intelectual à preservação do patrimônio digital, pois uma biblioteca pode ter o direito de acesso e uso da informação digital, mas não o de preservar.

Arquivos e bibliotecas nacionais. Fortalecer e instrumentalizar – em termos de equipamentos, metodologias e pessoal especializado – o Arquivo e a Biblioteca Nacional, para que possam desempenhar um papel ativo de liderança e de disseminadores de conhecimento na gestão da preservação dos documentos digitais. Alertar aos outros detentores de acervos digitais que a primeira linha de defesa contra a perda de informação de valor são eles – geradores, disseminadores e proprietários de informação digital.

Governo. Propor e estabelecer políticas públicas voltadas para a salvaguarda do patrimônio digital do país, incluindo alocação de recursos, estabelecimento de diretrizes e ordenamentos e disponibilização de instrumentos de divulgação para a conscientização da sociedade sobre os problemas decorrentes da fragilidade dos objetos digitais.

Recursos humanos. Incluir nos currículos das escolas de arquivologia, biblioteconomia e tecnologia da informação disciplinas que contemplem as questões de preservação digital; organizar e promover treinamento, cursos de extensão e especialização para as equipes graduadas atuantes na área.

Alianças e cooperação. É preciso lembrar que garantir o acesso à informação digital é um compromisso profundo e duradouro, cujo prazo é indefinido, e do qual tomam parte muitos atores. É fundamental, portan-

to, estabelecer pactos de cooperação entre governo, editores, indústrias de TI, bibliotecas, arquivos, museus, universidades, institutos de pesquisa etc. Sem alianças e cooperação, os programas mais amplos não serão absolutamente possíveis.

Acesso. Por fim, o mais importante: tudo isso é necessário para garantir acesso aos estoques de informações digitais de valor contínuo às suas respectivas comunidades-alvo, e, não menos importante, garantir também que essa informação – enquanto válida – permaneça compreensível para esta comunidade.

Notas

[1] LUKESH, Susan S. E-mail and Potential Loss to Future Archives and Scholarship or The Dog that Didn't Bark. *First Monday, Peer-Reviewed Journal on the Internet*, v. 4, n.9. September 6th 1999. Disponível em: <http://www.firstmonday.org/issues/issue4_9/lukesh/index.html>. Acesso em: 14/03/2005

[2] COOK, Terry. It's 10 o'clock: do you know where your data are? *Technology Review*, p. 48-53, January 1995. Disponível em: <http://www.cpsr.org/foia/PROFS_CASE/CanadianExperience>. Acesso em: 02/03/2003

[3] WARNER, Dorothy; BUSCHMAN, John. Studying the Reader/Researcher Without the Artifact: Digital Problems in the Future History of Books. *Library Philosophy and Practice*, v. 7, n. 1 (Fall 2004). Disponível em: <http://www.webpages.uidaho.edu/~mbolin/warner-buschman.htm>. Acesso em: 14/05/2005.

[4] LESK, Michael. Preservation of New Technology. Disponível em: <http://www.lesk.com/mlesk/cpa/cpa2.html>. Acesso em: 14/03/2005.

[5] CHEN, Su-Shing. The paradox of digital preservation. *Computer*, p.2-6, March 2001. Disponível em: <http://www.gseis.ucla.edu/us-interpares/pdf/ParadoxOfDigitalPreservation.pdf>. Acesso em: 14/03/2005

[6] CONSELHO NACIONAL DE ARQUIVOS, CÂMARA TÉCNICA DE DOCUMENTOS ELETRÔNICOS – CONARQ/CTDE. *Carta para a preservação do patrimônio arquivistico digital*. Rio de Janeiro : CONARQ, UNESCO, 2005, 21p. Disponível em <http://www.arquivonacional.gov.br/conarq/cam_tec_doc_ele/download/CartaPreservPatrimArqDigital-Conarq-2004.PDF>. Acesso em 06/06/05

[7] MARCONDES, Carlos Henrique; SAYÃO, Luis Fernando. The SciELO Brazilian Scientific Journal Gateway and Open Archives: A Report on the Development of the SciELO-Open Archives Data Provider Server. *D-Lib Magazine*, v.9, n.3, March 2003. Disponível em: <http://www.dlib.org/dlib/march03/marcondes/03marcondes.html>. Acesso em: 14/03/2005.

[8] MARCONDES, Carlos Henrique ; SAYÃO, Luis Fernando . Brazilian Digital Library of Theses and Dissertations. *International Information And Library Review Elsevier*, v. 35, n. 2-4, p. 265-279, 2003.

[9] CONSULTATIVE COMMITEE FOR SPACE DATA SYSTEM (CCSDS). *Reference model for Open Archival Information System (OAIS)*: recommendation. Washington, 2002. 139p. Disponível em <http://www.ccsds.org/documents/650x0b1.pdf >. Acesso em 02/02/2004.

[10] LEE, Kyong-Ho e outros. The State of the Art and Practice in Digital Preservation. *Journal of Research of the National Institute of Standards*

and Technology, v.107, n.1, p. 93-106, January-February 2002. Disponível em: <http://nvl.nist.gov/pub/nistpubs/jres/107/1/j71lee.pdf.>. Acesso em: 14/03/2005

[11] BULLOCK, Alison. *Preservation of digital information*: issues and current status. Ottawa : National Library of Canada, April 22, 1999. Disponível em: http://www.nic-bnc.ca/publications/1/p1-259-e.html. Acesso em: 11/03/2004.

[12] CONWAY, Paul. *Preservação no universo digital*. Rio de Janeiro: Projeto Conservação Preventiva em Bibliotecas e Arquivos, 2001, 34p.

[13] THOMAZ, Kátia. *A preservação de documentos eletrônicos de caráter arquivístico*: novos desafios, velhos problemas. 2004. Tese (Doutorado) – Escola de Ciência da Informação, UFMG, Belo Horizonte, 2004.

[14] TASK FORCE ON THE ARCHIVING DIGITAL INFORMATION. *Preservation digital information*: report of the Task Force on Archiving Digital Information. Washington, DC. : Commission on Preservation and Access, 1996. Disponível em: <http://www.rlg.org/ArchTF>. Acesso em: 02/02/2003.

[15] DAY, Michael. Metadata for digital preservation: an update. *Ariadne*. v.22. Disponível em: <http://www.ariadne.ac.uk/issue22/metadata/>. Acesso em: 09/05/2003

[16] ROTHENBERG, Jeff. Ensuring the longevity of the digital documents. *Scientific American*, p.24-29, January 1995. Disponível em: <http://www.clir.org/pubs/archives/ensuring.pdf >. Acesso em: 02/02/2003

[17] MEALLING, M. *The Network Solutions Personal Internet Name (PIN): A URN Namespace for People and Organizations. Network* Working Group Request for Comments: 3043. Diponível em: <http://www.ietf.org/rfc/rfc3043.txt>. Acesso em: 14/03/2005.

[18] WEIBEL, Stuart L. *PURLs*: Persistent Uniform Resource Locators. Dublin:OCLC. Disponível em: <http://purl.oclc.org/docs/new_purl_summary.html>. Acesso em: 14/03/2005.

[19] Digital Object Identifier System. Home Page: <http://www.doi.org/>. Acesso em: 14/03/2005.

[20] DURANTI, Luciana. *Preserving Authentic Electronic Art Over The Long-Term: The InterPARES 2 Project*. Disponível em: <http://aic.stanford.edu/sg/emg/pdfs/Duranti-EMG2004.pdf >. Acesso em 05/03/2003.

[21] LAVOIE, Brian. *Meeting the challenges of digital preservation*: the OAIS reference model. OCLC –Newsletter, n.243, p.26-30, Jan./Feb. 2000. Disponível em: <http://www.OCLC.org/research/publications/archive/2000/lavoie/>. Acesso em 05/03/2003.

[22] HD-Rosetta Archival Preservation Services. Home Page: <http://www.norsam.com/hdrosetta.htm>. Acesso em: 14/03/2005

[23] HEDSTROM, Margaret. Digital preservation: a time bomb for digital libraries. *Computer and the Humanities*, v.31, n.3, p.189-202, 1997/1998. Disponível em: <http://www.uky.edu/~kiernan/DL/hedstrom.html>. Acesso em 20/09/2003.

[24] HEDSTROM, Margaret. *Digital preservation: problems and prospects*. Disponível em: <http://www.dl.ulis.ac.jp/DLjournal/No_20/1-hedstrom/1-hedstrom.html>. Acesso em: 14/03/2005

[25] LORIST, Jeroen. Standard for digital libraries and archives: digital longevity. Betade Internal Reports, March 2001. Disponível em: <http://www.betade.tudelft.nl/reports/Lorist_StandardsLongevity_20010307.pdf>. Acesso em 14/03/2005

[26] ROTHENBERG, Jeff. Avoiding technological quicksand: finding a viable technical technical foundation for digital preservation. Disponível em: <http://www.clir.org/pubs/reports/rothenberg/contents.html>. Acesso em: 14/03/2005

[27] SHEPARD, Thom. Universal Preservation Format. *Abbey Newsletter*, v.22, n.5, 1998. Disponível em: <http://www.palimpset.stanford.edu/byorg/abbey/an/an22-5/an22-504.html>. Acesso em:02/02/2004.

[28] MARCONDES, Carlos Henrique; SAYÃO, Luis Fernando. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. *Ci. Inf., Brasília*, v. 30, n. 3, p. 24-33, set./dez. 2001. Disponível em:

<http://www.google.com.br/url?sa=U&start=2&q=http://www.ibict.br/cienciadainformacao/include/getdoc.php% 3Fid%3D514%26article%3D224%26mode%3Dpdf&e=7620>. Acesso em: 14/03/2005.

[29] The continuity of digital heritage. *Gateways*, n.61, February 2003. Disponível em: <http://www.nla.gov.au/ntwkpubs/gw/61/p04a01.html>. Acesso em: 14/03/2005.

Ferramentas de *software* livre para bibliotecas digitais

*Hélio Kuramoto**

1. Introdução

Uma das grandes dificuldades encontradas por bibliotecas e centros de documentação, em especial no Brasil, para disseminar os seus acervos, relaciona-se à identificação de uma ferramenta adequada para tratá-los, organizá-los e disseminá-los. Qualquer iniciativa para organizar, tratar e disseminar informação, seja referencial, seja em texto integral, exigia, até alguns anos atrás, um investimento relativamente importante, devido ao alto custo das soluções de *software*. As soluções eram dependentes de pacotes de *software* de gestão de bancos de dados, os quais eram, em sua maioria, comerciais e embutiam soluções proprietárias[1]. Em passado recente e mesmo hoje, muitas bibliotecas se deparam com dificuldades para automatizar o seu acervo bibliográfico, apesar de existirem diversas soluções comerciais e algumas de domínio público[2].

Recentemente, com o surgimento das novas tecnologias da informação agregadas ao movimento do *software open source*[3], também chamado *software* livre, várias alternativas de pacotes de *software* para construção e gestão de bibliotecas digitais têm sido distribuídos na *Web*. Verifica-se, dessa maneira, que os esforços para a construção e gestão de

*Doutor em ciências da informação e da comunicação pela Université Lumière – Lyon 2, França. Coordenador geral de projetos especiais e do Programa Biblioteca Digital Brasileira, do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). kuramoto@ibict.br

bibliotecas digitais são bem menores, sem necessidade de grandes investimentos em termos de infra-estrutura tecnológica. Essa facilidade será mostrada ao longo deste capítulo, mas antes, para contextualizar, será feita uma rápida discussão sobre o conceito de bibliotecas digitais e terminologias afins. Em seguida, será apresentada a iniciativa de arquivos abertos (Open Archives Initiative), os seus ideais e padrões. Essa iniciativa constitui um marco importante no desenvolvimento das bibliotecas digitais. Em seguida, serão apresentados alguns requisitos desejáveis para um pacote de *software* de administração de bibliotecas digitais, além de uma tipologia desses pacotes e uma breve descrição de alguns dos pacotes de *software*, *open source*, identificados na *Web*.

2. Bibliotecas digitais: contexto e definições

Até alguns anos atrás, década de 80, utilizavam-se para a disseminação da informação as bases de dados bibliográficas ou referenciais. Essas bases lidavam apenas com informações referenciais, pois não havia tecnologias que facilitassem o tratamento e a disseminação de conteúdos integrais. O custo das unidades de armazenamento e de memória era muito alto.

Com a popularização dos microcomputadores, o barateamento das unidades de armazenamento, das memórias e o surgimento das novas tecnologias da informação, começaram a surgir as bases de dados *full text*. Concomitantemente a esse fenômeno surgiram termos como bibliotecas eletrônicas, bibliotecas digitais, bibliotecas virtuais e até bibliotecas do futuro e bibliotecas sem paredes.

Segundo a Association of Research Libraries (ARL), existem várias definições para bibliotecas digitais, assim como, existem termos como biblioteca eletrônica e biblioteca virtual que são sempre utilizadas como sinônimos. Essa associação extraiu, de um trabalho[4] realizado por Drabenstott, os elementos comuns identificados nessas definições:

- a biblioteca digital não é uma simples entidade;
- a biblioteca digital requer tecnologias para interconectar os recursos de outras bibliotecas digitais;
- a interoperabilidade entre várias bibliotecas digitais e serviços de informação é transparente aos usuários finais;

- a meta principal é o acesso universal a bibliotecas digitais e serviços de informação;
- as bibliotecas digitais não se limitam a referências bibliográficas ou informações referenciais: elas se estendem aos artefatos digitais que não podem ser representados ou distribuídos em formato impresso.

A ARL tenta, portanto, caracterizar o que seja uma biblioteca digital, sem, no entanto, conceituá-la.

No início de um encontro do WG (DLib Working Group on Digital Library), foi proposta a seguinte definição:

Biblioteca digital é definida como sendo um conjunto de serviços apoiados por uma coleção de objetos informacionais que suporta usuários interessados nesses objetos, assim como organiza e preserva esses objetos disponíveis direta ou indiretamente por meio eletrônico ou digital.

A biblioteca digital não é meramente equivalente a uma coleção digitalizada com ferramentas de gestão de informação. Trata-se, também, de uma série de atividades que integram coleções, serviços e pessoas em suporte do completo ciclo de criação, disseminação, uso e preservação de dados, informação e conhecimento.

Verifica-se na literatura da área que os termos bibliotecas digitais, bibliotecas virtuais, bibliotecas eletrônicas designam um mesmo tipo de sistema de informação, portanto são termos sinônimos, apesar de cada um desses termos privilegiar determinado aspecto da tecnologia, assim:

- o termo bibliotecas eletrônicas enfatiza o fato de os acervos encontrarem-se registrados em meio eletrônico;
- bibliotecas digitais enfatizam o fato de a informação encontrar-se codificada em base digital;
- biblioteca virtual enfatiza o fato de o sistema utilizar tecnologias de realidade virtual. Alguns autores entendem também que nesse tipo de biblioteca a informação não está presente no servidor onde se encontra na biblioteca, mas em outros servidores remotos. Portanto, esse tipo de biblioteca encerra um conjunto de *links*.

Para efeito deste capítulo, o termo bibliotecas digitais designa uma aplicação envolvendo o registro, o tratamento e a disseminação de infor-

mações referenciais, assim como conteúdos em texto integral em meio digital.

3. A iniciativa dos *open archives* e as bibliotecas digitais

A iniciativa de *open archives* constitui um marco na história do desenvolvimento da comunicação científica e, por conseguinte, da publicação eletrônica e das bibliotecas digitais. Essa iniciativa promoveu a elaboração e o estabelecimento de padrões e protocolos para interoperar bibliotecas digitais.

Em julho de 1999, Paul Ginsparg, Rick Luce e Herbert Van de Sompel fizeram uma chamada para participação de uma reunião para explorar a cooperação entre arquivos de *e-prints* científicos. O encontro foi realizado em outubro de 1999, em Santa Fé, e foi denominado originalmente como a Reunião sobre Serviço Universal de *Preprint*, para o estabelecimento da Open Archives initiative (OAI) [www.openarchives.org]. O objetivo da OAI é contribuir de forma concreta para a transformação da comunicação científica. O veículo proposto para essa transformação é a definição de aspectos técnicos e organizacionais de uma estrutura para publicação científica na qual ambas as camadas, livres e comerciais, podem ser estabelecidas.

Assim, além dos aspectos técnicos, essa iniciativa deu origem a um movimento chamado *Open Archives and Open Access to Knowledge and Information in Science and Humanities*. Isso foi possível devido aos ideais e conceitos estabelecidos pelo Open Archives Initiative (OAI), que podem ser sintetizados nos seguintes termos: uso de *software* livre, também chamado *open source*, auto-arquivamento[5], criação de repositórios[6] de livre acesso, tanto institucionais quanto temáticos, uso de padrões de preservação de objetos digitais.

Em termos organizacionais, a iniciativa de arquivos abertos preconiza dois grandes conjuntos de atores[7], os provedores de dados (*data providers*) e os provedores de serviços (*service providers*).

Os provedores de dados são os gestores de arquivos de *e-prints* ou também denominados repositórios, aos quais são submetidos, diretamente pelos seus autores, os *pre-prints*, relatórios técnicos, dentre outros tipos de documentos. Os provedores de dados tornam disponíveis,

à comunidade, mecanismos para submissão ou auto-arquivamento, responsabilizam-se pelo armazenamento a longo prazo dos trabalhos, preocupam-se com a preservação dos objetos digitais e tornam disponíveis, para coleta, os metadados relativos aos trabalhos ou documentos publicados no seu repositório.

Os provedores de serviços, que também podem se caracterizar como agregadores, são instituições que se preocupam em coletar os metadados armazenados nos vários provedores de dados, oferecendo uma interface ou mecanismo de busca que possibilita aos usuários finais a consulta ou busca em um repositório que integra os metadados coletados a partir de diferentes provedores de dados.

Esse modelo de organização é bastante interessante, pois viabiliza e facilita o acesso à informação pela comunidade científica e tecnológica. Os provedores de serviços têm o papel de integrar diferentes provedores de dados, os quais armazenam informação em áreas do conhecimento específicas. A existência dos provedores de serviços temáticos facilita ao usuário final o acesso à informação especializada. No Brasil, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) vem trabalhando com essa perspectiva, assumindo o papel de um grande provedor de serviço, ou agregador, e promovendo a instituição de provedores de dados em diversas áreas do conhecimento, assim como a instituição de repositórios institucionais, facilitando, assim, o cumprimento de sua missão como órgão nacional de informação em ciência e tecnologia.

Para facilitar a exposição de metadados por parte dos provedores de dados e a sua captura por parte dos provedores de serviços, a OAI estabeleceu um conjunto de padrões: definiu-se o *Dublin Core*, sem os qualificadores, como padrão de metadados; o protocolo OAI-PMH (*Open Archive Initiative – Protocol of Metadata Harvesting*) como o mecanismo para interoperar os vários provedores de dados e os provedores de serviços. O OAI-PMH é um protocolo concebido e executado no contexto de um outro protocolo, o HTTP – *Hypertext Transfer Protocol*, protocolo utilizado pela *Web* para navegação em suas páginas.

Graças ao estabelecimento dessa iniciativa, vários pacotes de *software* para a construção e gestão de bibliotecas digitais surgiram, utilizando os ideais e padrões preconizados por essa iniciativa. Na seção 5.1 serão descritos alguns desses pacotes.

4. Requisitos desejáveis para um *software* de biblioteca digital

Existem inúmeros pacotes de *software* para construção e administração de uma biblioteca digital. Esses pacotes vão desde um *software* totalmente genérico como um SGBD[8], a partir do qual, aliado a um conjunto de programas, pode-se construir qualquer tipo de biblioteca digital, até pacotes que são específicos para determinado tipo de coleção como um sistema de gestão de eventos científicos, como o *Open Conference System*. A escolha de pacotes específicos como o *Open Conferece System* ou o *Open Journal System*, com propósitos específicos, demanda do usuário menor esforço, dado que esses pacotes de *software* são preparados para serem parametrizados, cabendo ao usuário apenas adequá-lo à sua necessidade. Esses pacotes de *software* não exigem esforço de programação e são simples de serem manipulados.

São os seguintes os requisitos básicos desejáveis para pacotes de *software* para gestão de bibliotecas digitais:

Portabilidade

A portabilidade é a habilidade de um pacote de *software* poder ser executado em diferentes ambiente operacionais. Poucos pacotes de *software* possuem essa habilidade. Hoje, essa característica seria muito desejável, tendo em vista que vivemos em um mundo heterogêneo, no qual convivem dois mundos, o dos *software open source*, ou *software* livre, e o dos *software* proprietários. Em outras palavras, vivencia-se com o mundo Linux e com o *Microsoft* simultaneamente.

Flexibilidade quanto à definição do padrão de metadados

Esse requisito indica se o pacote de *software* possui a habilidade de oferecer, ao usuário, outros padrões de metadados, além daquele que ele adota. Existem pacotes de *software* como o Nou-Rau e o Phronesis, que não possuem essa flexibilidade, pois eles implementaram um formulário de entrada de documentos baseado no padrão *Dublin Core*. Outros, porém, são mais flexíveis, como, por exemplo, o *Greenstone*, que dá ao usuário a possibilidade de criar o seu próprio padrão de metadados, apesar de ter o *Dublin Core* como padrão *default*. A importância de se ter essa

flexibilidade advém do fato de que nem sempre o *Dublin Core* é capaz de descrever adequadamente um determinado objeto digital.

Uso de padrões de interoperabilidade

Trata-se de uma característica também importante, pois o fato de um pacote de *software* ter disponível um padrão como o Z39.50 ou o protocolo OAI-PMH proporcionará, à biblioteca digital criada, a possibilidade de se integrar a outras iniciativas análogas, facilitando a criação de redes de bibliotecas digitais. O uso, por exemplo, do OAI-PMH permite a gestão de federação de bibliotecas digitais. Um exemplo dessa federação é a Biblioteca Digital de Teses e Dissertações, mantida pelo IBICT. Essa facilidade permite maior visibilidade e integração aos conteúdos de bibliotecas digitais.

Uso de linguagem de marcação XML

A linguagem de marcação XML tornou-se um padrão na *Web*, e o seu uso proporciona maior interoperabilidade entre os sistemas de bibliotecas digitais. Se nos anos 80 e 90 os sistemas de informação bibliográficos / documentários utilizavam a família de formato MARC para facilitar o intercâmbio de dados, hoje o XML ocupa esse espaço para a troca de dados e navegação na *Web*.

Capacidade de tratamento de múltiplos formatos de documentos

As novas tecnologias da informação propiciaram o surgimento de vários tipos de formatos para armazenamento de textos, imagem e som. Assim, um pacote de *software* para administrar uma biblioteca digital deveria ter facilidades para o tratamento de diversos formatos como o PDF, RTF, HTML, XML, DOC, GIF, JPEG e outros.

Possuir interfaces ergonômicas e adaptativas

Considerando que o público usuário não é normalmente especializado no manuseio de computadores e de linguagens de programação, é desejável que as interfaces oferecidas pelos pacotes de *software* de administração de bibliotecas digitais utilizem padrões de ergonomia com vistas a facilitar o seu uso por parte dos usuários menos especializados. Nesse contexto, não se deve esquecer a comunidade usuária portadora

de deficiência física. Assim, com o propósito de inserção social e digital, a ergonomia das interfaces é um aspecto importante a se considerar na escolha de pacotes de *software* para administração de bibliotecas digitais. Deve-se verificar também a existência de facilidades para adaptação das interfaces às necessidades estéticas das bibliotecas digitais.

Facilidades para estruturação dos documentos

Determinadas coleções necessitam de facilidades que proporcionem uma estruturação hierárquica e que, inclusive, os procedimentos de indexação do conteúdo considerem essa estruturação. Um exemplo desse tipo de necessidade é o caso de uma coleção de revistas. Existem duas alternativas para a sua estruturação: 1) considerar a revista como sendo uma unidade de informação, composta por várias subunidades representadas por cada um dos artigos contidos na revista; 2) considerar cada artigo que a revista contém como sendo uma unidade de informação.

Configuração dos procedimentos de indexação

Essa funcionalidade dá ao administrador da biblioteca digital a possibilidade de definir os metadados que se tornarão os campos de busca, ou pontos de acesso, aos registros de uma biblioteca digital. Alguns pacotes de *software* oferecem essa funcionalidade e dão a possibilidade de escolher não apenas os metadados que se deseja indexar, mas também se o conteúdo integral do documento deverá ser ou não indexado. O *Greenstone* ofecere essa possibilidade, oferecendo parâmetros para que se possa indicar se o conteúdo do documento deve ser indexado, e até, a indexação de seções ou capítulos em índices separados do conteúdo integral do documento. Dentre os pacotes de *software* que serão apresentados, o *Greenstone* é o único a oferecer essa facilidade. Essa característica repercute na interface de busca, a qual o ferece ao usuário a possibilidade de escolha dos metadados que poderão compor uma estratégia de busca. Portanto, a existência dessa funcionalidade não é apenas desejável como característica, mas imprescindível para oferecer ao usuário maior flexibilidade na construção de sua estratégia de busca.

Possui módulo de formatação de relatórios ou saídas

Nem sempre os pacotes de *software* de construção e administração de bibliotecas digitais oferecem facilidades para se formatar e gerar rela-

tório ou para se configurar a apresentação de documentos. Trata-se de uma funcionalidade muito desejável, pois eventualmente o usuário tem necessidades de imprimir o resultado de uma busca. Em alguns pacotes de *software*, como o Greenstone, o usuário pode configurar a forma de apresentação por meio do uso da linguagem HTML.

Foram detalhadas nesta seção as principais características desejáveis em um pacote de *software* para administração de bibliotecas digitais. Nem todos os pacotes de *software* possuem todas essas características, são poucos aqueles que possuem todas as facilidades apresentadas. A seguir, será apresentada uma tipologia de pacotes de *software* para administração de bibliotecas digitais.

5. Tipologia de pacotes de *software* para administração de bibliotecas digitais

Conforme mencionado anteriormente, existem vários pacotes de *software* para administração de bibliotecas digitais e cada um apresenta funcionalidades específicas, mas nem todas atendem às necessidades dos usuários que pretendem construir e manter uma biblioteca digital.

As soluções de *software*, para se criar uma biblioteca digital e mantê-la, vão desde o completo desenvolvimento de um sistema a partir de um SGBD (Sistema de Gestão de Bases de Dados) como Oracle, DB II, dentre outros, até pacotes de *software* desenvolvidos para essa finalidade.

O total desenvolvimento de um sistema para administrar uma biblioteca digital nem sempre é uma solução viável, tendo em vista o alto custo de desenvolvimento e a dependência de recursos humanos especializados tanto em informática quanto em informação com experiência nesse tema. Apesar de os pacotes de *software* nem sempre terem todas as funcionalidades desejáveis, ainda assim algumas delas poderão vir a ser úteis, principalmente quando esse pacote é um *software open source*, devido à possibilidade de se inserirem modificações de forma a atender às necessidades específicas de cada aplicação.

Os pacotes de *software* para construção e administração de bibliotecas digitais podem ser divididos nos seguintes tipos: 1) *software* de caráter genérico; 2) *software* de caráter específico.

Pacotes de *software* de caráter genérico

Os pacotes de *software* de caráter genérico são aqueles que podem ser utilizados para qualquer tipo de aplicação, podem ser utilizados na construção de qualquer tipo de biblioteca digital. Isto significa que esses pacotes de *software* são capazes de tratar, organizar, registrar e disseminar qualquer tipo de documentos. Por exemplo, o pacote de *software* NouRau pode ser utilizado para tratar, organizar, registrar e disseminar qualquer tipo de coleção de documentos, uma vez que ele utiliza um padrão de metadados compatível com o padrão Dublin-Core, que é genérico. Sabe-se, à priori, que esse padrão não consegue descrever alguns tipos de coleções, mas dependendo das exigências do usuário, ele poderá ser adotado em uma gama variada de coleções.

A seguir, será feita uma breve descrição de cada um desses pacotes de *software*. Devido à variedade de pacotes de *software* existentes, não se pretende ser exaustivo enumerando todos eles, mas será feita uma breve descrição do *software Greenstone*, e em seguida, será apresentado um quadro (Quadro 1) com os pacotes de *software* conhecidos pelo autor e suas características técnicas.

Greenstone

Esse pacote de *software* foi desenvolvido pela University of Waikato, da Nova Zelândia, no contexto do projeto New Zeland Digital Library (www.nzdl.org). O Greenstone (www.greenstone.org) oferece uma nova forma de tratar, registrar e disseminar informação na Internet. Trata-se de um *software open source*, distribuído mediante licença[9] GPL. A versão atualmente em distribuição é a 2.53. Concomitantemente a essa versão, os desenvolvedores desse *software* estão distribuindo um novo Greenstone, todo redesenhado, denominado Greenstone 3.0. A distribuição do Greenstone conta com apoio da Unesco. Seguem algumas das funcionalidades desse *software*:

Ambientes operacionais – *software* multiplataforma, executável tanto em ambiente UNIX compatível, quanto em ambiente Windows e Mac OS X. Desenvolvido em linguagem C++ e PERL, utiliza utilitários, todos *software open source*, como o MG (Managing Giga), biblioteca de funções que facilita a indexação e recuperação de informação, o SGBD. Portanto, atende ao requisito de portabilidade.

Formatos de documentos suportados " O Greenstone, ao ser descarregado do seu portal, vem com *plugins*[10] desenvolvidos para tratar determinados tipos de formatos, como por exemplo:

TEXTPlug – trata documentos com extensão .txt;
HTMLPlug – trata documentos com extensão htm ou html
PDFPlug – trata documentos com extensão PDF
RTFPlug – trata documentos com extensão RTF
WordPlug – trata documentos com extensão DOC

Esses *plugins* são fornecidos originalmente pelo próprio Greenstone, mas nada impede que o usuário possa desenvolver um *plugin* específico para tratar um determinado tipo de documento com características específicas definidas pelo próprio usuário. O *software* oferece essa facilidade.

Formato de armazenamento interno – O Greenstone, ao armazenar o conteúdo de um documento, converte-o para um formato compatível com o XML.

Padrão de metadados – Originalmente, o Greenstone vem com a definição do padrão Dublin Core. No entanto, como ele é todo parametrizável, pode-se definir um outro padrão de metadados, que pode vir inserido no documento, no caso de documentos em HTML, ou em um arquivo em formato XML, denominado metadata.xml.

Modalidades de funcionamento – O Greenstone oferece três modalidades de funcionamento: ambiente local; ambiente Internet; ou em CD-ROM. Isso significa que o usuário poderá criar uma biblioteca digital em seu micro, sem que ela seja acessível por um outro computador. Da mesma forma, o usuário poderá criar uma biblioteca digital, utilizando outra modalidade de instalação, e torná-la disponível para acesso via *Web*, assim como exportá-la para um disco CD-ROM, facilitando a sua instalação em outro computador.

Forma de operação – A configuração de uma coleção no Greenstone se dá por meio de parametrização, cujos parâmetros são armazenados em um arquivo do tipo TXT, denominado CONFIG.CFG. Recentemente, as últimas versões desse *software* tornaram disponível uma interface gráfica que facilita, àqueles não iniciados em programação, a parametrização de uma aplicação de bibliotecas digitais, de uma forma mais ergonômica e dirigida. Trata-se de uma interface amigável e fácil de ser manipulada

por pessoas não especializadas. Essa interface é denominada *Greenstone Librarian Interface* (GLI).

Padrões de interoperabilidade – O *Greenstone* em suas últimas versões vem com dois padrões: o Z39.50, para busca distribuída, e o OAI-PMH (*Open Archives Initiative – Protocol of Metadata Harvesting*), protocolo que permite a coleta de metadados.

Formas de submissão de documentos – A entrada de documentos nos repositórios criados e administrados pelo *Greenstone* pode ser realizada de três maneiras: por FTP[11], por meio de um diretório, ou pasta e por meio do protocolo HTTP. Até a presente versão, esse *software* não oferece a facilidade de auto-arquivamento, ou seja, não existe a funcionalidade de o usuário submeter, *on-line*, um determinado arquivo contendo o documento (*upload*).

Facilidades de apresentação e recuperação de informação " O *software Greenstone* oferece uma funcionalidade que raramente é oferecida por outros pacotes. Trata-se do parâmetro *Classify*, que permite a criação de listas de documentos ordenados por título, autor ou qualquer outro metadado definido pelo administrador da biblioteca digital. Essa facilidade permite ao usuário da biblioteca digital consultá-la por meio de navegação ou visualização em listas classificadas. Essas listas podem ser tanto ordenadas alfabeticamente, quanto segundo uma determinada hierarquia. Além disso, esse *software* permite que se especifique a forma de apresentação dos resultados de uma busca ou consulta em linha.

Idioma das interfaces – As interfaces do Greenstone podem ser visualizadas em vários idiomas, inclusive o português do Brasil.

Pacotes de *software* de caráter específico

Ao contrário dos pacotes de *sofwtare* descritos na seção precedente, os pacotes de *software* de carácter específico são aqueles que foram desenvolvidos para atender a determinados tipos de aplicações. Um exemplo desse tipo de pacote de *software* é o *Open Journal Systems (OJS)*. Esse *software* foi desenvolvido para atender à demanda de construção e administração de uma revista científica eletrônica. O IBICT customizou esse *software* para a língua portuguesa do Brasil e o distribui com a denominação de Sistema Eletrônico de Editoração de Revistas (SEER). Hoje, o SEER é utilizado por mais de 15 (quinze) revistas científicas que estão disponíveis na *Web* (http://www.ibict.br/secao.php?cat=SEER/Revistas).

Um outro exemplo de *software* nessa categoria é o TEDE – Sistema de Publicação Eletrônica de Teses e Dissertações, desenvolvido e distribuído pelo IBICT.

No quadro 2 será apresentada uma lista de alguns desses *software* com algumas de suas características técnicas. Da mesma forma que a descrição da categoria anterior, não será feita uma lista exaustiva de pacotes de *software* nessa categoria, tendo em vista a enorme variedade desses pacotes e a exigüidade deste espaço. Será apresentado, portanto, um pequeno conjunto de *software*.

Considerando que o termo bibliotecas digitais designa um sistema de informação que trata tanto de informações referenciais quanto de conteúdos integrais, não poderia deixar de apresentar pacotes de *software open source* para automação de bibliotecas, detalhados no quadro 3.

Quadro 1 – Relação de Pacotes de *Software* de Caráter Genérico (*Open Source*)

Características Técnicas	Pacotes de *Software* para Bibliotecas Digitais					
	CDSware[12]	E-Prints[13]	Nou-Rau[14]	Phronesis[15]	Fedora[16]	DSpace[17]
Ambiente Operacional	Unix / Linux	Unix / Linux	Unix / Linux	Unix / Linux	Unix / Linux / Windows	Unix / Linux
Tecnologias Utilizadas	Python, Apache	Perl, MG (Managing Gigs)	PHP, Perl, ht//Dig, Apache	Gcc, Perl, Java, Apache	Java (J2SDK)	Java, Tomcat, Apache, Ant
Software de banco de dados	MySQL	MySQL	PostgreSQL	-	MySQL, Oracle 9i ou McKoi	PostgreSQL 7.3+
Padrões de Interoperabilidade	OAI compatível	OAI-PMH		Z39.50	OAI-PMH	OAI-PMH, Web Services
Metadados	Dublin Core, Marc21	Dublin Core, sem qualificadores	Dublin Core compatível, sem qualificadores	Dublin Core compatível, sem qualificadores	Dublin Core	Dublin Core
Facilidades de importação/exportação	-	-	-	-	Sim	Sim
Documentação	Informações no *site*.	Suporte Técnico via *mailing list*[18]	Informações no *site*.	Informações no *site*,	Documentação no *site*	Documentação no *site*
Aplicações	Servidor de Documentos Eletrônicos	Repositórios de uma forma geral.	Repositórios de uma forma geral	Repositórios de uma forma geral	Repositórios de uma forma geral	Repositórios de uma forma geral
Tipos de Formatos de documentos aceitos	Sem restrições	DOC, PDF, HTML, TXT, MP3, JPEG	Configurável pelo administrador	TXT,DOC,RTF, HTML, PDF, JPEG, MP3	Não há restrições	Não há restrições

Quadro 2 – Relação de Pacotes de *Software* de Caráter Genérico (*Open Source*)

Características Técnicas	Pacotes de *Software* para Bibliotecas Digitais			
	OJS[19, 20]	OCS[21, 22]	TEDE[23]	iVia[24]
Ambiente Operacional	Unix	Unix	Unix / Windows	Unix
Tecnologias Utilizadas	PHP, Apache	PHP, Apache	PHP	C++
Software de banco de dados	MySQL	MySQL	MySQL	MySQL, Berkeley DB
Padrões de Interoperabilidade	OAI-PMH	OAI-PMH	OAI compatível	OAI-PMH
Metadados	Dublin Core, sem qualificadores	Dublin Core, sem qualificadores	MTD-Br	MARC, Dublin Core,
Facilidades de importação/exportação	-	-	-	
Documentação	Suporte Técnico por meio de um fórum cujo endereço encontra-se no seu *site*..	Suporte Técnico por meio de fórum cujo endereço encontra-se em seu *site*.	Documentação no *site*	Documentação no *site*
Aplicações	Gestão e Publicação Eletrônica de Revistas Científicas na web.	Gestão e Organização de Conferências e publicação dos anais na web.	Publicação Eletrônica de Teses e Dissertações	Construção de Bibliotecas Virtuais, permitindo a criação automática e manual de coleções de *links* e de conteúdos em texto integral.
Tipos de Formatos de documentos aceitos	DOC, HTML, PDF	DOC, HTML, PDF	PDF, HTML	Não há restrições

Quadro 3 – Relação de Pacotes de *Software* de Caráter Específio (*Open Source*)

Características Técnicas	Pacotes de *Software* para Bibliotecas Digitais				
	GNUTECA[25]	KOHA[26]	PHPMyLibrary[27]	OpenBiblio[28]	PMB[29]
Ambiente Operacional	Linux	Unix / Linux	Originalmente desenvolvido para Windows / Linux	Unix / Linux / Windows	Unix Linux
Tecnologias Utilizadas		PHP, Perl, Apache	PHP, Apache	PHP, Apache	PHP, Apache
Software de banco de dados	PostgreSQL	MySQL, PostgreSQL ou Oracle	MySQL	MySQL	MySQL
Padrões de Interoperabilidade	ISO2709	-	-	-	-
Metadados	Marc21	MARC	Segue estritamente o USMARC	Unimarc	Unimarc
Facilidades de importação/exportação	Sim	-	-	-	-
Documentação	Documentação no *site*.	Documentação no *site*.	Documentação no *site*.	Documentação no *site*	Documentação no *site*
Aplicações	Automação de bibliotecas com as seguintes funções: OPAC[30], catalogação, aquisição e Circulação.	Automação de bibliotecas com as seguintes funções: OPAC, aquisição, catalogação e circulação.	Automação de bibliotecas com as seguintes funções: OPAC, aquisição, catalogação e circulação.	Automação de bibliotecas com as seguintes funções: OPAC, catalogação e circulação, além de funções administrativas.	Automação de bibliotecas, com as seguintes funções: OPAC, aquisição, catalogação e circulação, controle de periódicos, arquivo de autoridades.

6. Considerações finais

Foram apresentados, neste capítulo, alguns pacotes de *software* com suas caracerísticas principais. É bom ressaltar que esses pacotes não são os únicos existentes. Ao contrário, são apenas alguns dos quais o autor do capítulo pôde experimentar ou que faziam parte do seu conhecimento, resultado de estudos de prospecção tecnológica realizados no laboratório de tecnologias da informação do IBICT.

Todos os pacotes apresentados têm uma característica comum, são *software open source*, portanto, *software* livre, e quase todos funcionam apenas em ambiente Linux compatível. Existem, é claro, outras alternativas de *software*, tanto livres quanto comerciais. O leitor poderá identificá-las na *Web* utilizando mecanismos de busca como Google, Miner, AltaVista ou outro qualquer.

Um outro *site* muito interessante e que traz sempre muitas novidades em termos de tecnologias da informação para a construção de bibliotecas digitais é o D-Lib Magazine, cuja url é: http://www.dlib.org/.

Notas

[1] Soluções proprietárias são soluções comerciais cujos códigos fontes não são entregues, mas apenas o código objeto.

[2] Domínio público são pacotes de *software* disponibilizados ao público, livre de qualquer custo ou taxas.

[3] Entende-se por *software* open source aquele cuja distribuição acompanha o seu código fonte. Isso significa que o usuário tem além do código de máquina do *software*, o seu código fonte, o que lhe dá condições para alterar e adequar o *software* segundo as suas necessidades e eventualmente distribuí-lo. Normalmente esses pacotes de *software* são livres de custo, mas eventualmente eles podem ter um custo associado, segundo os serviços que são oferecidos vinculados ao *software*.

[4] DRAENSTOTT, Karen M. Analytical review of the library of the future, Washington, DC: Council Libary Ressources, 1994.

[5] Os próprios autores arquivam os seus trabalhos nos repositórios eletrônicos.

[6] Entende-se por repositórios, bancos de dados contendo conteúdos completos e respectivos metadados.

[7] Segundo [Van de Sompel, Herbert & Lagoze, Carl] participaram da Convenção de Santa Fé dois tipos de atores:

data providers e service providers, os quais são defnidos como: "A

data provider is the manager of an e-print archive, acting on behalf of the authors submitting documents to the archive.

As pointed out above, the data provider of an open archive will, at least, provide a submission mechanism, a long-term storage system and a mechanism that enables third parties to collect data from the archive; A service provider is a third party, creating end-user services based on data stored in e-print archives. For instance, a service provider could implement a search engine for mathematical e-prints stored in archives worldwide".

[8] SGBD – Sistema de Gestão de Bases de Dados; em inglês DBMS – Database Management System

[9] A licença GPL é uma forma de distribuição de *software* livre.

[10] Plugin é uma espécie de módulo de programa que executa um tipo específico de processamento.

[11] FTP é a sigla de File Tranfer Protocol, trata-se de um padrão para transferência de arquivos na Internet.

[12] Desenvolvido por CERN. (http://cdsware.cern.ch/*download*)

[13] Desenvolvido pela Southampton University. (http://software.eprints.org)

[14] Desenvolvido pelo Instituto Vale do futuro em parceria com a Unicamp. (http://www.rau-tu.unicamp.br/nou-rau/)

[15] Desenvolvido pela Universidad de Monterrey. (http://copernico.mty.itesm.mx/phronesis/project/phronesis14.html)

[16] Desenvolvido pela University Virginia Library e Cornell University. (http://www.fedora.info/)

[17] Desenvolvido pela MIT Libraries em parceria com a Hewlett-Packard Company. (http://www.dspace.org/)

[18] http://sofwatre.eprints.org/maillist.php

[19] Desenvolvido pela University of British Columbia. (http://pkp.ubc.ca/ojs)

[20] O IBICT customizou esse *software* para a língua portuguesa em colaboração com a University of British Columbia e o distribui sob a denominação de SEER

[21] Desenvolvido pela University of British Columbia. (http://pkp.ubc.ca/ocs)

[22] O Centro de Infomações Nucleares, da CNEN, em colaboração com o IBICT, customizou o *software* OCS para a língua portuguesa e o IBICT o distribui sob a denominação de SOAC.

[23] Desenvolvido pelo IBICT, com apoio da FINEP. (http://www.ibict.br/)

[24] Desenvolvido pela biblioteca da University of California, Riverside. (http://infomine.ucr.edu/iVia/)

[25] Desenvolvido por.... hhttp://www.gnuteca.org.br/

[26] Desenvolvido por Horowhenua Library Trust and Katipo Communications (http://www.koha.org/)

[27] Desenvolvido nas Filipinas (http://www.phpmylibrary.sourceforge.net/)

[28] http://obiblio.sourceforge.net/

[29] http://www.pizz.net/

[30] OPAC significa On-line Public Access Catalog

Referências

1 SOMPEL, Herbert van de; LAGOZE, Carl. The Santa Fé Convention of the Open Archives Iniciative. *D-Lib Magazine*, Virgínia, v. 6, n. 2, 2000. Disponível em: <http://www.dlib.org/dlib/february00/ vandesompel-oai/02vandesompel-oai.htm.> Acesso em: 11 mar. 2005.

2 BOLLACKER, Kurt D.; LAWRENCE, Steve; GILES, C. Lee. *CiteSeer:* an autonomous web agent for automatic retrieval and identification of interesting publications. SmealSearch. Disponível em: <http://smealsearch2. psu. edu/cache/papers/Business/594/

http:zSzzSzwww.csd.uch.grzSz~markatoszSzpaperszSzACM98. Autonomous.Agents.CiteSeer.pdf/DID.pdf/>. Acesso em: 21mar. 2005.

3 BRIN, Sergey; PAGE, Lawrence. *The anatomy of a large-scale hypertextual web search engine*: CiteSeer. Disponível em: <http://citeseer.ist.psu.edu/brin98anatomy.html>. Acesso em: 21 mar. 2005.

4 O'NEILL, EdwardT.; LAVOIE, Brian F.; BENNETT, Rick. Trends in the evolution of the public web. *D-Lib Magazine*, Virginia, v. 9, n. 4, Apr. 2003.

Disponível em:<http://www.dlib.org/dlib/april03/lavoie/04lavoie.html>. Acesso em: 12 mar. 2005.

5 CHRISTIAN, Eliot J. GILS: What is it? Where's it going? *D-Lib Magazine*, Virginia, Dec. 1996. Disponível em:

<http://www.dlib.org/dlib/ december96/12christian.html>. Acesso em: 21 mar. 2005.

6 STAPLES, Thorton; WAYLAND, Ross; PAYETTE, Sandra. The Fedora Project: an open-source digital object repository management system. *D-Lib Magazine*, Virginia, v. 9, n. 4, Apr. 2003. Disponível em: <http://www.dlib.org/dlib/april03/staples/04staples.html>. Acesso em: 12 mar. 2005.

7 SMITH, MacKenzie e outros. DSpace: an open source dynamic digital repository. *D-Lib Magazine*, Virginia, v. 9, n.1, Jan. 2003. Disponível em: <http://www.dlib.org/dlib/january03/smith/01smith.html>. Acesso em: 12 mar. 2005.

8 TRISKA, Ricardo; CAFÉ, Lígia. Arquivos abertos: subprojeto da Biblioteca Digital Brasileira. *Ciência da Informação*, Brasília, v. 30, n. 3, set./dec. 2001.

Disponível em: <http://www.ibict.br/cienciadainformacao/viewarticle.php?id=241&layout=abstra>. Acesso em: 21 mar. 2005.

9 WILLINSKY, John. Proposing a knowledge exchange model for scholarly publishing. *Current Issues in Education*, v. 3, n. 6, 2000. Disponível em: <http://cie.ed.asu.edu/volume3/number6/>. Acesso em: 23 mar. 2005.

10 _____; WOLFSON, Larry. The indexing of scholarly journals: a tipping point for publishing reform? *Journal of Electronic Publishing*, v. 7, n. 2, 2001. Disponível em:<http://www.press.umich.edu/jep/07-02/willinsky.html>. Acesso em: 22 mar. 2005.

3. Dimensões de uso

O novo papel das tecnologias digitais na comunicação científica

*Sely M. S. Costa**

1. Introdução

Como tenho dito em outros textos, a comunicação científica constitui um dos tópicos que têm sido muito explorados e discutidos na ciência da informação, ao longo das últimas quatro décadas. Por esse motivo, tem contribuído de forma significativa para a construção de conhecimento na área. Conseqüentemente, são vários os modelos teóricos, as abordagens e os contextos encontrados na literatura para seu estudo. Isso, por sua vez, reflete uma variedade de aspectos por meio dos quais o processo de comunicação entre pesquisadores tem sido estudado. A partir principalmente da década de 90, estudos sobre os impactos que a introdução de tecnologias da informação no ambiente acadêmico provoca na comunicação científica vêm contribuindo para o enriquecimento do debate a respeito das questões pertinentes ao tópico e que são, por conseguinte, relevantes para estudo.

Mais recentemente, observa-se um debate caloroso na literatura a respeito de como a comunicação entre pesquisadores tem refletido as mudanças provocadas pela introdução das tecnologias da informação no meio acadêmico. Uma dessas mudanças diz respeito às interações no

* PhD, CID/UnB, selmar@unb.br

seio das comunidades científicas, assim denominadas de forma genérica, visto que podem ser definidas em diferentes níveis. É importante destacar, por exemplo, que, em um nível mais abrangente, o conceito de comunidade científica inclui os pesquisadores, os bibliotecários, os provedores de acesso à Internet, os centros de computação, editores, agências de fomento etc. É nesse nível de abrangência, portanto, que se considera a questão das tecnologias digitais e os modelos de comunicação científica resultantes do seu uso, por pesquisadores acadêmicos, para comunicação de suas pesquisas.

Desse modo, o capítulo aborda cinco questões consideradas essenciais para a discussão do tema e apresenta um panorama sucinto sobre o que pensam alguns autores que as têm debatido na literatura recente. Primeiro, discute a questão da comunicação científica nos dias atuais, apresentando um modelo híbrido que ilustra o processo de comunicação científica em função da coexistência do meio impresso e do meio eletrônico para comunicação da pesquisa. Segundo, comenta alternativas que estão disponíveis aos pesquisadores para acesso à informação em termos de serviços baseados no meio eletrônico. Terceiro, discorre sucintamente sobre como o acesso aberto tem influenciado a discussão a respeito da acessibilidade da informação científica no seio das comunidades científicas, mais especificamente no que concerne às agências de fomento, às universidades e aos editores. Quarto, argumenta como uma espécie de "filosofia aberta" tem influenciado a comunicação científica hoje. Finalmente, apresenta resultados de pesquisa que mostra quais reflexos podem ser observados no Brasil, especialmente no que diz respeito às iniciativas com arquivos abertos no País.

2. Um modelo híbrido do processo de comunicação científica

Quando Garvey; Grifith (1979) apresentaram seu modelo do processo de comunicação, resultado de um estudo amplo com pesquisadores da psicologia, e com enfoque na questão temporal, certamente – como todo autor – gostariam que o trabalho tivesse um fator de impacto elevado. No entanto, podem não ter imaginado o impacto que o trabalho, de fato, teve no estudo da comunicação científica por estudiosos de outras

disciplinas e sobre pesquisadores das mais diversas áreas do conhecimento. A verdade é que o modelo de Garvey; Grifith, por seu aspecto seminal no estudo do processo, tem sido amplamente explorado. Serve, assim, de base para um sem-número de outros estudos.

É importante ressaltar a adaptabilidade do modelo com relação a novos fatores que, ao provocarem mudanças no processo de comunicação científica, requerem novas formas de representação do modelo de Garvey; Grifith. Assim, novas realidades, interpretadas por seus autores como mudanças no processo de comunicação científica, têm sido possíveis de ser representadas com base no modelo dos autores.

É o caso dos impactos que a introdução de tecnologias de informação no ambiente acadêmico provoca no processo de comunicação entre pesquisadores. Hurd (1996), ao estudar essa questão, em relação a cientistas naturais, propôs um modelo inteiramente baseado no meio eletrônico. Costa (1999) a estudou também em relação a cientistas sociais e propôs um modelo híbrido do processo de comunicação onde a coexistência dos meios impresso e eletrônico é ilustrada. Entretanto, podem-se destacar dois aspectos do modelo híbrido que lhe permitem se manter apropriado para ilustrar o processo, enquanto perdurar a referida coexistência dos dois meios: sua atualidade como forma de representação do processo de comunicação científica em diferentes períodos de tempo e a possibilidade de representar diferentemente o processo de comunicação entre pesquisadores de diferentes divisões do conhecimento.

No que concerne à questão de mudanças que ocorrem ao longo do tempo, a forma de representação do modelo permite ilustrar o processo em, por exemplo, quatro momentos distintos. Primeiro, a representação do processo desde os primeiros episódios do uso de comunicação eletrônica, quando se pode representar a presença pouco significativa do meio eletrônico em todas as suas etapas. Segundo, sua representação no período compreendido pela década de 90, em que há uma presença significativa do meio eletrônico, porém infinitamente maior na comunicação informal do que na formal. Terceiro, a representação do processo nos dias atuais, em que se observa uma representatividade um pouco mais equilibrada nos aspectos informais e formais do processo, embora ainda maior na comunicação informal, e o crescimento na formal relacionado com a publicação, embora mais especificamente em periódicos. Finalmente, a

representação no futuro, em que as perspectivas parecem ser de uma representatividade cada vez maior do meio eletrônico em todas as etapas do processo.

A figura 1 é uma tentativa de ilustrar o processo como observado no momento atual, procurando representar a dimensão da presença do impresso e do eletrônico em cada uma de suas etapas. É uma adaptação do trabalho de Costa (1999), o qual adaptou o trabalho de Hurd (1996), que, por sua vez, se baseou em Garvey; Grifith (1979). É importante notar que na adaptação de Hurd nem todas as etapas do modelo de Garvey; Grifith estão representadas, visto que a autora simplificou o modelo. Assim, o modelo aqui apresentado também representa uma simplificação do modelo dos autores. A representação reflete a interpretação desta autora, não levando em consideração nenhuma divisão específica do conhecimento. Na visão dos leitores, pode haver uma interpretação diferente, o que estimula o debate a respeito da representatividade do modelo híbrido para o processo de comunicação científica, levando em conta a coexistência do meio impresso e do meio eletrônico.

Apesar da dinamicidade do processo de comunicação científica, em função das mudanças que o uso de tecnologias de informação provoca, é importante ressaltar que duas questões tendem a permanecer inalteradas. A primeira diz respeito ao uso do meio eletrônico em maior escala nas etapas do processo concernentes à comunicação informal do que naquelas relacionadas à comunicação formal, em todas as áreas do conhecimento. A segunda se refere às diferenças disciplinares e a tendência, por exemplo, de maior presença do meio eletrônico na comunicação formal entre cientistas das áreas exatas e naturais do que entre cientistas sociais e humanistas.

É possível, no entanto, elaborar versões diferentes do modelo, dependendo da divisão do conhecimento sobre a qual se quer representar o processo de comunicação. Mais que isso, o modelo híbrido não somente reflete a coexistência dos dois meios na comunicação da pesquisa – coexistência essa que, por sua vez, permanecerá também por muito tempo –, mas permite, igualmente, registrar a complementaridade gradual e crescente do meio eletrônico em relação ao meio impresso. Não obstante, o aspecto crescente da complementaridade dos dois meios poderá se tornar substituição, pelo menos em relação a algumas das etapas do pro-

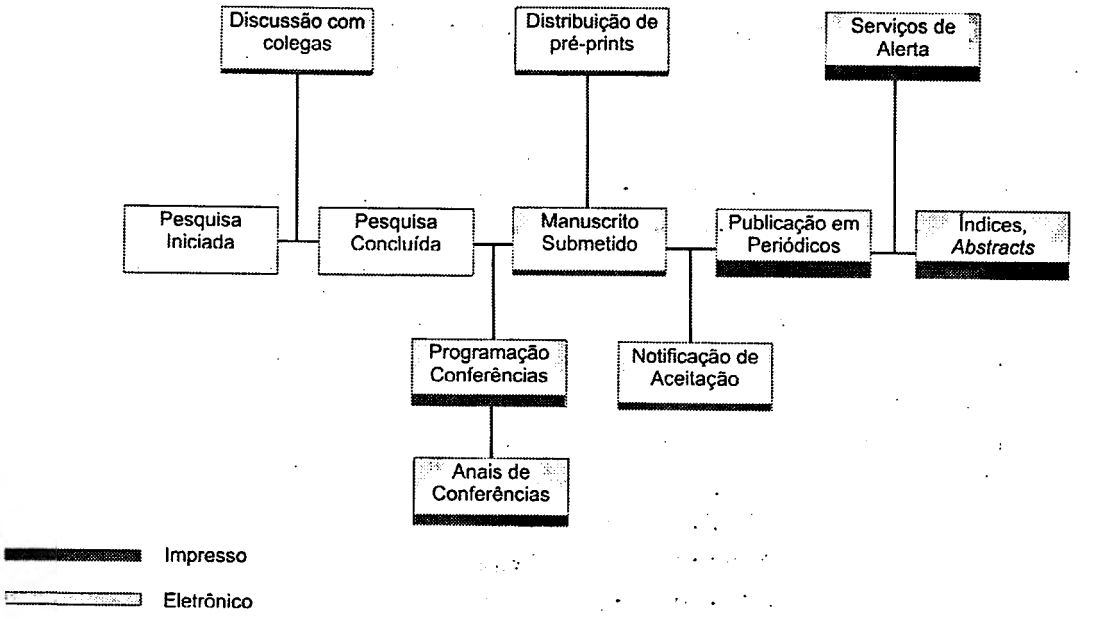


Figura 1: Modelo híbrido do processo de comunicação científica, adaptado de Costa (1999)

cesso para as quais o meio impresso vem apresentando representatividade cada vez menor, como é o caso da submissão de manuscritos a editores e da notificação de aceitação/rejeição de trabalhos para publicação.

Outros aspectos parecem tender a se perpetuar refletindo a coexistência dos dois meios. Entre eles, o registro formal do conhecimento em livros, periódicos e outros formatos, assim como os canais de acesso a eles. Em outras palavras, tanto a publicação da pesquisa quanto os meios de acesso aos conteúdos publicados parecem tender a refletir a coexistência do meio impresso e do meio eletrônico indefinidamente. Isso, por sua vez, tem reflexo nas bibliotecas como coleções onde fontes de informação estão organizadas e disponíveis, e nos serviços de informação pelos quais se tem acesso às fontes e seus conteúdos. Há um número cada vez maior de serviços de informação baseados no meio eletrônico, com vistas a tornar o acesso à informação rápido e diminuir o *gap* entre o momento do registro do conhecimento e o momento do acesso à informação. Parece oportuno, portanto, apresentar umas poucas considerações a respeito de alguns serviços de informação para pesquisadores científicos, disponíveis no meio eletrônico.

3. Rápidas considerações sobre serviços de informação para a comunidade acadêmica, baseados no meio eletrônico

É possível pensar que desde que existem computadores, existem, também, serviços de informação eletrônicos. Sem dúvida, os serviços de informação muito cedo se beneficiaram dos desenvolvimentos da computação. Mais ainda, da combinação da computação e das telecomunicações. A Internet, nesse contexto, vem contribuindo para tornar disponíveis cada vez mais, e rapidamente, conteúdos, em princípio acessíveis em nível global. Em princípio, porque a acessibilidade não depende apenas da disponibilidade, mas, e principalmente, do letramento tecnológico e da formação (conhecimento necessário ao entendimento dos conteúdos disponíveis) dos usuários da informação.

No meio acadêmico, uma série de questões tem provocado a adoção crescente pela comunicação eletrônica. Primeiro, a pressão dos pares, das agências de fomento e das próprias universidades (Costa, 2000). Segundo, o chamado determinismo tecnológico que tem mostrado a influência da disponibilidade sobre o uso de tecnologias[1] por mais que se reconheçam, na literatura, fatores outros – particularmente sociais – responsáveis pela adoção de tecnologias de informação (Gomes, 1999). Terceiro, o processo quase natural de adoção de inovações, estudado por Rogers (1995), em que, precoce ou tardiamente, como ressalta Borgman (2001), a maioria dos pesquisadores acadêmicos acaba por decidir pela adoção de tecnologias de informação para comunicação da pesquisa. Há outros, não discutidos aqui. O que se quer ressaltar é uma certa imposição sobre os pesquisadores, nos dias atuais, para que usem serviços de informação baseados no meio eletrônico. Estes, por outro lado, constituem um meio de aumentar o acesso à informação publicada, o que, por sua vez, aumenta a visibilidade dos pesquisadores.

De fato, a acessibilidade depende, em grande parte, do potencial de serviços de informação que propiciem uma busca rápida, dinâmica e eficiente. Como observa Lawrence (2001), ao discutir a questão da maior visibilidade de autores que têm seus trabalhos disponíveis *on-line*, com acesso livre,

> A disponibilidade *on-line* de um artigo pode não aumentar significativamente acesso e impacto, caso não haja serviços de busca eficientes e

abrangentes, visto que um percentual substancial da literatura precisa ser indexado por esses serviços antes que os cientistas o considerem útil.

Há, portanto, um número cada vez maior de produtos e serviços eletrônicos de informação, os quais vêm se tornando indispensáveis aos pesquisadores para o acesso à informação útil a suas pesquisas. Do ponto de vista da comunicação científica, serviços, seus produtos e os conteúdos que tornam disponíveis tendem a constituir um único *locus*, no sentido em que algumas formas de organização da informação digital permitem a criação de repositórios onde autores depositam os conteúdos por eles gerados, tornando-os imediatamente acessíveis a suas audiências. Um exemplo são os repositórios institucionais, comentados no item 5.

4. O acesso aberto e seu impacto nas universidades, nos editores comerciais e nas agências governamentais de fomento à pesquisa científica

Brody; Harnad (2004) têm, persistentemente, chamado atenção para o fato de que o acesso livre (*open access*) a resultados de pesquisa maximiza o acesso à pesquisa propriamente dita, maximizando e acelerando o impacto dessas pesquisas e, conseqüentemente, a produtividade, o progresso e as recompensas da pesquisa. Neste contexto, os autores observam que o impacto:

- fornece a medida do tamanho da contribuição que a pesquisa dá a futuras pesquisas (publique ou pereça);
- gera futuros financiamentos de pesquisa;
- contribui para a produtividade da pesquisa e para o apoio financeiro da instituição do pesquisador;
- avança a carreira do pesquisador;
- promove o progresso da pesquisa (grifos dos autores).

Em linha com o que observam Brody; Harnad, Lawrence (2001), apresenta resultados de uma pesquisa que mostrou um crescimento de 336%, em média, nas citações a artigos disponíveis *on-line*, em relação a artigos

publicados *off-line*, na mesma fonte. O autor ressalta que "para maximizar o impacto, minimizar a redundância e acelerar o progresso científico, autores e editores deveriam visar a tornar a pesquisa fácil de ser acessada". O acesso tende, então, a ser a questão crucial do progresso científico em qualquer área do conhecimento.

Nos últimos cinco a seis anos, tem-se identificado um movimento entre pesquisadores acadêmicos, principalmente das ciências exatas e naturais, no sentido de que seus resultados de pesquisa estejam disponíveis, gratuitamente, o mais amplamente possível. Comunidades científicas de disciplinas como a física, a matemática, a computação e a biologia quantitativa (arXiv: http://arxiv.org), a economia (RePEc: http://repec.org), as ciências da vida (PubMed Central: http://www.pubmedcentral.nih.gov), a medicina e biologia (PloS: http://www.plos.org) e as ciências cognitivas – psicologia, neurociência, lingüística, filosofia, biologia, entre outras disciplinas (CogPrints: http://cogprints.org), por exemplo, desenvolveram soluções amplamente conhecidas.

Tais iniciativas provocaram um amplo e acalorado debate na literatura mundial a respeito do novo modelo de publicação, em que "direito livre, irrevogável, mundial e perpétuo de acesso" a trabalhos publicados, como definido na Reunião de Bethesda de 11 de abril de 2003 (Bethesda Statement..., 2003) é reivindicado em nível global.

Suber (2003) chama atenção para o fato de que o sistema atual de publicação é contrário a esse etos. De fato, como ressaltam Chan; Costa (2005), editores comerciais têm atribuído preços excessivos e imposto barreiras de permissão sobre publicações de pesquisas que são amplamente financiadas com recursos públicos. No entanto, a informação científica e técnica é, fundamentalmente, um bem público global, que deve estar livremente disponível para o benefício de todos (Alberts, 2002). É compreensível, portanto, o debate que se observa na literatura sobre o tema. A *Nature*, por exemplo, tem estimulado essa discussão. Em data recente, a *Nature* publicou o resultado de uma enquete em que procurou identificar qual seria o desejo dos seus leitores para 2005. Após a conclusão da pesquisa, a própria revista decidiu se incluir nos resultados, apresentando seus desejos, chamados de "resoluções para os leitores da revista", e que incluem, entre outras idéias (*Wish list* 2005):

> Boa informática: Resolva visitar uma biblioteca real para folhear periódicos e livros em papel pelo menos uma vez em 2005. E lembre a você mesmo e aos estudantes que o conhecimento disponível sobre um assunto não é restrito às respostas a buscas no Google;
>
> Boa comunicação: Resolva ser bravo e compartilhe seu conhecimento científico com o público em geral [....] Convide um jornalista para almoçar a fim de se conhecerem mutuamente. Ofereça-se para falar a um grupo de alunos da educação básica, ou mesmo do jardim de infância [...];

O que se depreende da brincadeira da *Nature* (há outras três recomendações) com seu público é a importância que a visibilidade dos pesquisadores tem, a necessidade do acesso amplo à informação científica e a imposição de que a informação científica deve estar disponível a todos, por meio de serviços de informação diversificados e de qualidade.

Todo esse debate tem reflexo nas universidades, nas agências de fomento e no próprio governo. Em alguns países desenvolvidos, como a Inglaterra, essa questão tem sido objeto de deliberação por parte de governos, universidades e agências de fomento há mais de cinco anos. Nos países em desenvolvimento, nem tanto. O que tem chamado mais atenção são os reflexos que a questão do auto-arquivamento tem provocado.

Chan; Costa (2005) ressaltam as experiências dos governos americano e britânico com relação ao papel que governo e agências de fomento exercem nesse contexto e que reflexo isso tem sobre editores comerciais. Os autores comentam que, nos países ricos, esses organismos

> estão começando a questionar a eficiência do sistema de publicação atual, dominado por interesses comerciais, e estão começando a demandar que a pesquisa que eles financiam esteja tão amplamente acessível quanto possível.

No caso britânico, o *Wellcome Trust*, maior instituição privada de financiamento do Reino Unido, anunciou seu apoio às iniciativas com acesso livre em 2003. O Comitê de Ciência e Tecnologia da *House of Commons* publicou, em 20 de julho de 2004, um relatório resultado de oito meses de levantamentos sobre essa questão (acesso a publicações científicas) no País. Duas das inúmeras recomendações do relatório são destacadas no artigo de Chan; Costa:

O governo deve prover fundos para todas as universidades do Reino Unido criarem repositórios institucionais de acesso livre;

Os autores de artigos baseados em pesquisa financiada pelo governo devem depositar cópias deles em seus repositórios institucionais.

No caso americano, Chan; Costa destacam a decisão da House of Representatives de que o National Institute of Health (segunda maior instituição de financiamento para pesquisa nos Estados Unidos) desenvolvam uma política de requerer que cientistas financiados pelo instituto depositem seus artigos no *PubMedCentral*, quando aceitos para publicação em um periódico.

As decisões e recomendações dessas instituições e países são muitas e mostram a importância da questão do acesso livre à informação no seio da comunidade científica e da sociedade em geral. Têm, certamente, impacto mundial. Urge, portanto, que países em desenvolvimento, como o Brasil, realizem estudos e desenvolvam e implementem projetos que visem à inserção do País nesse contexto. Recai sobre as bibliotecas universitárias enorme desafio. Sobre as comunidades científicas da ciência da informação, desafio maior ainda, assim como da ciência da computação. Assim, é fundamental que se conheçam em profundidade, e se estudem nos cursos de formação dessas áreas, as questões relacionadas com o que se chama neste capítulo de Tríplice AA, como se discute a seguir, visando a dar uma noção genérica sobre os termos.

5. O tríplice AA: acesso aberto, arquivos abertos, auto-arquivamento: influência de uma "filosofia" aberta?

Uma análise dos trabalhos apresentados na 7ª International Conference on Electronic Publishing-ElPub2003 (Costa; Moreira, 2003) revelou que dentro da abordagem geral do tema da conferência (publicações eletrônicas), o ambiente acadêmico é identificado como o principal cenário tanto para pesquisa quanto para desenvolvimento, o que aponta para as comunidades científicas como seus principais atores. Pode-se, então, afirmar que à medida que o meio eletrônico, gradual e crescentemente, substitui

o meio impresso no contexto da publicação eletrônica, as mudanças na comunicação aumentam também, provocando crescimento e diversidade nas interações entre pesquisadores e no desenvolvimento do conhecimento. Isso, por seu turno, acelera o acesso à informação.

Da mesma forma, mudanças paradigmáticas relacionadas com uma necessária filosofia aberta estão progressivamente acontecendo, dando lugar a uma nova ordem mundial que, por sua vez, é subjacente às preocupações de pesquisadores em relação à publicação eletrônica. Uma sumarização das idéias discutidas nas sessões gerais da ElPub2003 mostrou, de fato, que a tendência principal da comunicação eletrônica hoje parece ser em direção a uma filosofia aberta, levantando questões sobre *software* aberto, acesso aberto, arquivos abertos e outras (Costa; Moreira, 2003).

Neste ponto, vale a pena ressaltar as explicações de Lagoze; Van de Sompel (2001) a respeito do termo arquivos abertos. Os autores observam que nesse termo, a palavra arquivo é usada para representar a idéia de um repositório onde se armazenam informações. Quanto à palavra aberto, o sentido é de uma interface de máquina aberta que facilite tornar conteúdos de diversos autores disponíveis, aumentando a visibilidade da produção de comunidades científicas.

Deusdará (2005) chama atenção para duas questões essenciais nessa discussão. A primeira diz respeito às afirmações de Unsworth, intimamente relacionadas com a abordagem de Lagoze; Van de Sompel e o guarda-chuva técnico da interoperabilidade (abordagem da multiplicidade de atores). A autora destaca, do trabalho de Unsworth (2004):

> As economias envolvidas nesses sistemas são três: uma economia de prestígio, primordial aos autores (produtores de conteúdo), importante mas secundária para os outros atores; uma economia monetária, primordial para os editores, não muito importante para os produtores de conteúdo na maioria dos casos, e importante, mas não realmente primordial, para as bibliotecas; e uma economia subsidiária, primordial para as bibliotecas que são financiadas pelas universidades como um bem público, e mais importante para os produtores de conteúdo do que eles geralmente imaginam (Unsworth *apud* Deusdará, 2005, p. 18).

A segunda, derivada do pensamento de Van de Sompel (2004), no qual se registra:

> Existe uma crescente insatisfação com o sistema de comunicação científica já estabelecido (...), resultado de vários fatores, incluindo o rápido aumento do preço das assinaturas, a questão dos direitos autorais, a lentidão entre o momento da obtenção do resultado e sua real publicação e restrições quanto ao que pode ser publicado e como isso pode ser disseminado (Van de Sompel *apud* Deusdará 2005, p. 19).

Na verdade, a questão do preço das assinaturas tem sido uma das principais, se não a principal motivação para os movimentos do acesso aberto e dos arquivos abertos. Prosser (2003) chama atenção para o fato de que por três séculos o modelo básico para a comunicação científica nas ciências exatas e naturais permaneceu imutável, com o periódico científico no papel central. Contudo, observa Prosser,

> A confluência dos preços crescentes dos editores, os orçamentos restritos das bibliotecas e as novas tecnologias eletrônicas de comunicação estão criando o ambiente certo para uma transformação marcante em como se compartilham novas informações dentro da comunidade desses cientistas.

Uma dessas transformações está relacionada com os repositórios institucionais como "coleções digitais que capturam e preservam a produção intelectual de uma universidade ou uma comunidade multiuniversitária". O autor compara as funções dos periódicos tradicionais, nomeadamente registro, certificação, *awareness* e arquivamento, com as dos repositórios institucionais, observando que nestes a certificação não é garantida, pelo menos nos mesmos moldes da "revisão independente e internacional, pelos pares". Uma possível solução, sugere Proser, seria uma rede de repositórios institucionais ligadas a periódicos referendados, caso em que, em razão do acesso livre aos repositórios institucionais, os periódicos teriam, igualmente, que prover acesso livre.

Nesse ponto as idéias de Prosser encontram eco no que Harnad e outros (2001) classificam como vias verde e dourada para o acesso aberto. A primeira é representada pelo auto-arquivamento por parte dos autores que obtêm permissão (sinal verde) dos editores de periódicos que aceitaram seus artigos para publicação, de depositarem seus trabalhos em um servidor de arquivos abertos com acesso livre. A segunda diz respeito aos peri-

ódicos científicos eletrônicos de acesso livre. Na primeira, portanto, o acesso livre é garantido pelos arquivos abertos quando estes tornam disponíveis, livremente[1], artigos publicados em periódicos científicos referendados. Na segunda, o acesso livre é garantido diretamente pelos próprios periódicos.

Observe-se que a abordagem de Harnad e outros traz para debaixo do "guarda-chuva técnico para interoperabilidade prática", de Lagoze; Van de Sompel (2004), os editores. Quanto aos repositórios institucionais, provocam a entrada das bibliotecas em cena. Todos certamente centrados na preocupação com a acessibilidade ampla e irrestrita à informação, resultado do movimento recente em direção à filosofia aberta para os arquivos em que os conteúdos gerados por pesquisadores estão disponíveis, os *softwares* utilizados na criação de repositórios de diversos tipos e o acesso, questão central em toda essa discussão.

6. Resultados de pesquisa recente sobre arquivos abertos no Brasil

Estudo recente (DEUSDARÁ, 2005) investigou como instituições acadêmicas brasileiras, tais como universidades e institutos de pesquisa, têm respondido ao movimento mundial não somente dos arquivos abertos, mas também de uma espécie de borbulhante filosofia aberta (Costa; Moreira, 2003). Neste sentido, tanto o número crescente de serviços e produtos que têm sido criados – particularmente para pesquisadores acadêmicos – quanto as mudanças que têm sido observadas no ambiente da pesquisas e nas comunidades científicas parecem ser resultado do acalorado debate encontrado na literatura, nos últimos três, quatro anos, sobre tópicos relacionados com a Iniciativa com Arquivos Abertos (*Open Archives Initiative – OAI*). Mais que isso, são certamente produto dessa espécie de filosofia aberta.

A preocupação com a questão dos arquivos abertos, periódicos eletrônicos, repositórios institucionais, bibliotecas digitais etc, aponta para a necessidade de pesquisas sobre o tópico, à medida que reflete novos paradigmas para a comunicação científica. Isso porque têm surgido tanto novos problemas de pesquisa sobre o sistema de comunicação científica, quanto novas formas de lidar com eles. Com essa motivação em mente, iniciou-se, em julho de 2004, um estudo *cross-sectional*, a ser realizado

regularmente pelos próximos dois ou três anos[2], com o objetivo de explorar as experiências com arquivos abertos no Brasil. Na primeira etapa, descrita nesta seção, o objetivo foi mapear as iniciativas brasileiras com implementações relacionadas com arquivos abertos.

Aspectos metodológicos do estudo

O estudo centrou o foco no uso do protocolo OAI-PMH (*Open Archives Initiative-Protocol for Metadata Harvesting*). Nesse contexto, considerou quatro tipos de iniciativas: o primeiro diz respeito ao envolvimento independente e direto de autores com o auto-arquivamento. O segundo envolve o desenvolvimento e a implementação de repositórios institucionais. O terceiro tipo se relaciona com iniciativas que visam a criar periódicos eletrônicos. Finalmente, o quarto tipo de iniciativa considerada no estudo diz respeito às que visam ao gerenciamento de conferências e publicação dos anais em formato eletrônico, em ambiente aberto.

Uma das suposições do estudo foi a de que o que vem sendo implementado no País em termos de OAI, envolvendo a comunidade científica, está de acordo com o que Lagoze; Van de Sompel (2001) chamaram de "guarda-chuva técnico para interoperabilidade prática" (Figura 2). Os autores discutem a multiplicidade de comunidades que exploram hoje a estrutura técnica da OAI para coleta de metadados (*metadata harvesting*), observando que uma variedade de atores pode agora ser identificada no contexto das comunidades de OAI-PMH, que incluem autores, editores, bibliotecários, agências de fomento etc. Isto é, comunidades científicas no contexto mais amplo do termo, conforme discutido no item 2 deste capítulo.

Uma segunda suposição para o estudo foi a de que a comunidade científica brasileira deve estar consciente do possível aumento da sua visibilidade, quando torna seus conteúdos disponíveis por meio de uma implementação baseada no OAI-PMH, interoperável, portanto. Como citado anteriormente, a pesquisa de Lawrence (2001) ressaltou a relação entre disponibilidade *on-line* de um trabalho e seu impacto. No entanto, quer com acesso livre ou não, uma implementação baseada na OAI constitui certamente um fator que contribui para aumentar a visibilidade do pesquisador.

Para a primeira etapa da pesquisa, a população investigada compreendeu 22 instituições acadêmicas apontadas pelo IBICT como envolvidas

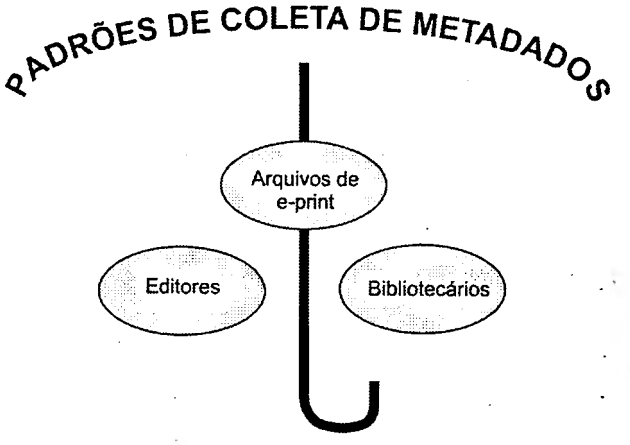


Figura 2: Estrutura da Iniciativa com Arquivos Abertos para Múltiplas Comunidades. Fonte: Deusdará, 2005, adaptado de Lagoze; Van de Sompel, 2001.

com OAI no Brasil. Das 22 instituições apontadas, 19 foram contatadas por *e-mail* e três por telefone, visando a uma consulta mais aprofundada sobre o trabalho com arquivos abertos realizado na instituição. Nove instituições responderam à primeira consulta (seis por *e-mail*), mas somente quatro confirmaram ter implementado alguma iniciativa com arquivos abertos: o Ibict, o Inpe, a UnB e a UCB.

Uma entrevista estruturada foi realizada com pessoas das quatro instituições, resultando na identificação de seis iniciativas, visto que tanto na UnB quanto no Inpe foram identificadas duas unidades administrativas distintas trabalhando com aplicações envolvendo OAI-PMH. O roteiro da entrevista consistiu de quatro grupos de questões, incluindo: características do respondente (cuja intenção foi unicamente a de deixá-lo(a) à vontade); tipo de implementação e suas características em termos de *software* usado, tipo de aplicação, *raison d'être* etc; percepções do respondente sobre vantagens e desvantagens da OAI; e expectativas sobre o futuro da OAI no País.

Síntese dos resultados do estudo

Características dos respondentes:

Três respondentes são do sexo masculino e uma do sexo feminino. Três têm formação em áreas das ciências sociais e humanas e um das ciências exatas e naturais. Dois têm doutorado, um mestrado e um bacharelado. Os quatro são responsáveis pela implementação da OAI em suas instituições.

Tipos de implementação e suas características:

Foi identificado o uso dos quatro tipos de iniciativas com arquivos abertos,definidos no estudo, quais sejam: auto-arquivamento em qualquer repositório, repositórios institucionais, periódicos eletrônicos e anais de conferências.

Todas as iniciativas identificadas usam o protocolo OAI-PMH, independentemente do *software* adotado. Todas se baseiam em *softwares* livres.

As iniciativas com arquivos abertos no País são lideradas por pesquisadores (autores), editores e bibliotecários, confirmando o modelo de interoperabilidade de Lagoze; Van de Sompel.

Os *softwares* utilizados e suas finalidades são:

- o ePrint, para auto-arquivamento, por parte do Ibict, em uma aplicação que envolve o depósito, pelos autores, de artigos referendados para publicação em periódicos, trabalhos apresentados em eventos e relatórios de pesquisa;
- o TEDE, para repositório institucional da Universidade Católica de Brasília, para o depósito de teses e dissertações;
- o URLib Service, desenvolvido pelo Instituto Nacional de Pesquisas Espaciais para dois tipos de aplicações: repositório institucional envolvendo o depósito de artigos de periódicos, capítulos de livros, livros relatórios de pesquisa e outros documentos; anais de conferências.
- o SEER (tradução para o português do Open Journal System, da Universidade de British Columbia, Canadá), para a criação de periódicos eletrônicos. Vem sendo usado pelo Ibict e pelo Centro de Formação de Recursos Humanos em Transportes (Ceftru), da Universidade de Brasília.
- o SOAC (tradução para o português do Open Conference System, da Universidade de British Columbia, Canadá) foi usado pelo Departamento de Ciência da Informação da UnB para a organização de duas conferências internacionais.

Percepção dos respondentes

Na percepção dos respondentes da pesquisa, o principal objetivo da

implementação de iniciativas com arquivos abertos é ampliar a visibilidade dos autores e suas organizações, assim como agilizar o processo de publicação de trabalhos científicos. A perspectiva para o futuro, na opinião deles, é de que haja um crescimento rápido na adoção de arquivos abertos no País.

7. Considerações finais

Parece pertinente considerar o guarda-chuva técnico para interoperabilidade prática como uma espécie de novo modelo do processo de comunicação científica no que concerne à representação dos aspectos relacionados com os atores do processo. Neste sentido, o modelo destaca a inclusão gradual de novos atores do processo, levando-se em conta que as mudanças quanto a essa questão se iniciaram pela ação de pesquisadores na qualidade de autores e usuários da informação científica. A adesão gradual de editores e das bibliotecas, como também de agências de fomento, é uma questão inexorável e irreversível. O que aponta para novas mudanças é o modelo que envolve o pagamento para disseminação (pagar para publicar, objeto de discussão em outros trabalhos). O modelo representa, no entanto, a ampliação das fronteiras de uma igualmente nova comunidade científica formada por adotantes, precoces ou tardios, das inovações que arquivos abertos, periódicos eletrônicos e outras formas de comunicação eletrônica representam.

Notas

[1] É importante lembrar que o conceito de arquivo aberto não inclui o acesso livre, como destacado por Van de Sompel e outros autores.

[2] Apesar de a pesquisa estar planejada para ser realizada, repetidamente, nos próximos anos, não constitui um estudo longitudinal, visto que não envolverá o mesmo grupo de sujeitos sendo entrevistado ao longo desse período.

Referências

ALBERTS, B. *Engaging in a worldwide transformation*: our responsibility as scientists for the provision of global public goods, President's Address to the Fellows of the National Academy of Sciences, USA, Washington D.C., 29 April 2002.

BETHESDA Statement on Open Access Publishing. 2003. Disponível em: http://www.earlham.edu/~peters/fos/bethesda.htm. Acesso em 23/-06/2004.

BORGMAN, Christine L. A premissa e a promessa de uma infra-estrutura global de informação. *Revista de Biblioteconomia de Brasília*, v. 25, n. 1, p. 91-126, jan/jun. 2001.

BRODY; Tim; HARNAD, Stevan.*The research impact cycle*. Disponível em: http://opcit.eprints.org/feb19oa/harnad-cycle.ppt. Acesso em 17 set. 2004.

CHAN, Leslie; COSTA, Sely. Participation in the global knowledge commons: challenges and opportunities for research dissemination in developing countries. New Library World, v. 106, n. 3/4(1210/1211), p. 141-163, 2005.

COSTA, Sely M. S. de. Mudanças no processo de comunicação científica: o impacto do uso de novas tecnologias. In: MUELLER, SUZANA P.M.; Passos, Edilenice (Org.). *Comunicação científica*. Brasília: CID/UnB, 2000. P. 85-106. (Estudos avançados em Ciência da Informação, 1).

COSTA, Sely M. S. de. *The impact of computer usage on scholarly communication amongst academic social scientists*. Loughborough, 1999. 291p. (Doctoral Thesis) – Department of Information Science, Loughborough University.

COSTA, Sely M. S.; MOREIRA, Ana Cristina S. The diversity of trends, experiences and approaches in electronic publishing: evidences of a paradigm shift on communication. In: COSTA, S. M. S.; CARVALHO, J.A.C.; BAPTISTA, A.A.; MOREIRA, A.C.S. From information to knowledge: proceedings of the 7th ICCC/IFIP International Conference on Electronic Publishing. Guimarães: Universidade do Minho, 2003, p. 5-9.

DEUSDARÁ, Ana Paula Oliveira. *Estudo sobre iniciativas com arquivos abertos no Brasil*. Brasília, 2005. (Monografia de Graduação). CID/Universidade de Brasília.

GOMES, Suely Henrique de Aquino. Inovação tecnológica no sistema formal de comunicação cientifica : os periódicos eletrônicos nas atividades de pesquisa dos acadêmicos de cursos de pós-graduação brasileiros. Brasília, 1999. (Tese de Doutorado) CID/Universidade de Brasília.

GARVEY, William D.; GRIFFITH, Berver G. Communication and information processing within scientific disciplines: empirical findings for Psychology. In: GARVEY, William D. *Communication*: the essence of science. Oxford: Pergamon Press, 1979. p. 127-147.

HARNAD, Stevan et al. *The access/impact problem and the green and gold roads to open access*. 2001. Disponível em: http://www.ecs.soton.ac.uk/~harnad/Temp/impact.html, acesso em 18 jul. 2004.

HARNAD, Stevan et al.. Mandated online RAE CVs linked to University Eprint Archives.Ariadne, n 35, 2003.

HURD, Julie M. Models of scientific communication systems. In: Crawford, Suzan Y.; Hurd, Julie M; Willer, Ann C. *From print to electronic*: the transformation of scientific communication. Medford, NJ: Information Today, 1996. p. 9-33.

LAGOZE, Carl; VAN DE SOMPEL, Herbert. The open archives initiative: building a low-barrier interoperability framework. 2001. Disponível em: http://www.openarchives.org/documents/jcdl2001-oai.pdf, acesso em 12 mar.2003.

LAWRENCE, Steve. Free online availability substantially increases a paper's impact. *Nature webdebates*. Disponível em: http://www.nature.com/nature/debates/e-access/Articles/lawrence.html, acesso em 20 de novembro de 2004.

PROSSER, David. Information revolution: can institutional repositories and open access transform scholarly communications? *The ELS Gazette*, v. 15, jul. 2003. Disponível em: http://www.the-elso-gazette.org/magazines/issue/features/features1.asp, acesso em 17 ago. 2004.

ROGERS, E. M. *Diffusion of innovations*. 4. Ed. New York: Free Press, 1995.

SUBER, Peter. Removing the barriers to research: an introduction to open access for librarians. College & Research Libraries News, 64: 92-94, 2003.

UNSWORTH, John. Open access, open archives and open source in higher education. 2004. Disponível em: http://www3.isrl.uiuc.edu/~unsworth/siue.2-04.html, acesso em 16 fev. 2005.

VAN DE SOMPEL, Herbert. Rethinking scholarly communication: building the system that scholars deserve. *D-Lib Magazine*, v. 10, n. 9, set. 2004. Disponível em http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html, aceso em 21 out. 2004.

Wish list 2005. *Nature*, v. 432, n. 7020, 23/30 dez. 2004. Disponível em: http://www.nature.com/nature, acesso em 13 fev. 2005.

A interface do usuário e as bibliotecas digitais

*Sueli Mara Soares Pinto Ferreira**
*Patrícia Cristina do Nascimento Souto***

1. Evolução do conceito de interface

Ressaltar a importância da interface quando se analisa qualidade de bibliotecas digitais é hoje condição *sine qua non*. No entanto, a interface nem sempre foi considerada importante, pelo contrário, uma análise de sua evolução demonstra que já foi o componente do sistema de menor relevância, por conta de dois fatores fundamentais: a limitação das máquinas em termos de capacidade de processamento e no fato dos sistemas serem utilizados por um pequeno grupo de pessoas com alto grau de especialização.

O desenvolvimento tecnológico, a extensão de seu uso a cada vez maior número de pessoas com diversificados interesses e níveis de conhecimento, sua inserção no cotidiano e em todas as atividades humanas e a evolução cognitiva ocorrida nos indivíduos no uso destas ferramentas tecnológicas, formam um ambiente propício para a constante reformulação deste conceito. As interfaces deixam de ser baseadas em

* Coordenadora da Portcom/Rede de Informação em Ciências da Comunicação dos países de língua portuguesa e do Núcleo de Pesquisa "Design de Sistemas Virtuais Centrado no Usuário". Professora doutora do Departamento de Biblioteconomia e Documentação da ECA/USP. smferrei@usp.br
** Pesquisadora da Portcom e do Núcleo de Pesquisa "Design de Sistemas Virtuais Centrado no Usuário", aluna de mestrado do Programa de Pós-Graduação em Ciências da Comunicação, área de concentração em ciência da informação e documentação. patriciacn@uol.com.br

linhas de comando e textos e começam a adicionar elementos da linguagem visual, resultando em interfaces gráficas[1], que ampliam as possibilidades e a qualidade do acesso, manipulação e comunicação entre homem e computador.

Neste sentido, interface passa a ser, de acordo com Moran (1981) "a parte de um sistema computacional utilizada pelo usuário para interagir física, perceptiva e conceitualmente com o próprio". Esta definição amplia a noção do conceito de interface, determinando claramente a presença de dois componentes complementares:

- componentes físicos – são os recursos tangíveis compostos por dispositivos de *hardware* com os quais os usuários realizam as atividades motoras e perceptivas (teclado, mouse, teclado, tela, impressora, ícones, comandos, *menus*, ferramentas, dentre outros) e, dispositivos de *software* que possibilitam que o usuário interaja com o sistema, por meio da implementação dos processos computacionais necessários para controle dos dispositivos de *hardware*; e
- componentes cognitivos – são recursos abstratos desenvolvidos pelo sistema a partir de diretrizes, conceitos e valores pré-definidos, para priorizar os processos cognitivos voltados ao raciocínio, à interpretação e ao processamento de informações do usuário durante a realização de suas tarefas.

Descarte a presença dos componentes cognitivos, o conceito ainda mantém seu foco e objeto de preocupação no sistema como um fim em si mesmo.

Atualmente, o conceito de interface agrega propostas e contribuições de várias áreas do conhecimento (como engenharia cognitiva, ciência da informação, psicologia e outras) e passa a centralizar sua preocupação no processo de interação homem-máquina e homem-homem através da máquina. Entende-se *interação* como a relação entre usuários (categorias, expertise e sistemas), conteúdos (tipos, níveis e relevância) e sistemas (interface, funcionalidades básicas e complementares) (Tsakonas, Kapidakis, Papatheodorou, 2004).

Segundo Beaudouin (2004) "...projetar a interação ao invés de desenhar a interface significa mudança do foco em busca da qualidade da

interação entre o usuário e o computador: a interface de usuário é o meio, não o fim".

Com esta proposta, a interface passa a ser percebida tanto como o *meio* para a interação usuário-sistema, quanto como uma *ferramenta* que oferece os instrumentos para este processo comunicativo.

Considerar a interface como o instrumento para construção da interação entre sistema e usuário apresenta uma série de desafios, sendo os principais:

(a) como conhecer, estudar e apreender conhecimento sobre os modelos mentais que usuários têm em relação as suas interações com o mundo;
(b) como articular, explorar e representar os modelos mentais identificados junto aos usuários em um projeto conceitual coerente e que se mantenha continuamente alinhado as suas necessidades, contexto, evoluções e movimento ao longo do tempo e do espaço;
(c) como acompanhar e considerar as evoluções nos modelos mentais dos usuários;
(d) como alcançar e gerenciar a efetividade e naturalidade na interação entre homem (modelo mental) e o sistema (interface representando o modelo conceitual proposto) (Horgen, 2001).

O conceito de interface continua evoluindo e atingindo novas gerações que discutem cada vez mais a aplicação de inteligência artificial[2] para promover interfaces mais adaptativas[3], sensíveis ao entorno e focada na tarefa e no modelo mental do usuário, formando com isso o que se denomina de "interfaces inteligentes". Esta nova geração de interfaces visa tornar a interação do usuário com a máquina mais eficiente (completar tarefas com menor esforço), mais efetiva (fazer a coisa adequada na hora certa e apoio ao processo de construção do conhecimento) e mais natural (suporte a linguagem natural) (Horgen, 2001).

Entretanto, criar e desenvolver uma interface com esse papel demanda que sua operacionalização seja constantemente alinhada às definições, objetivos, estratégias e diretrizes delineadas na concepção do sistema. De outra forma, a evolução conceitual sobre interface se desconecta do seu verdadeiro propósito que é alcançar melhores níveis de produtividade do usuário, configurando-se num desalinhamento potencialmente gerador de baixos níveis de adoção e uso de sistemas de informação.

2. A interface como um dos fatores que influenciam a adoção das bibliotecas digitais

Os fatores que afetam a aceitação dos usuários e, conseqüentemente, a adoção do sistema, segundo o "Modelo de Adoção de Tecnologia" (TAM) de Thong e outros (2002), são determinados pelas percepções de sua *utilidade* e *facilidade* de uso.

Facilidade de uso identifica a percepção de que inexiste esforço por parte do usuário para manusear o sistema. Quanto mais fácil for a interação do usuário com o sistema, mais ele sentirá a utilidade do mesmo e crescerá sua intenção de adotá-lo.

Utilidade determina a percepção de que o sistema aumenta a performance do usuário durante o desempenho de suas tarefas. Isto é determinante na adoção do sistema, pois, segundo os autores do modelo acima, usuários são mais inclinados a adotá-los se eles podem proporcionar funções de maior valor.

Com base neste modelo de aceitação (TAM), seus autores iniciaram estudos e pesquisas para identificar os fatores de adoção em bibliotecas digitais. Após um estudo de dois anos com mais de dois mil estudantes/usuários da *Electronic Library (E-Library) da Open University of Hong* (OUHK), Thong e outros (2004, 2002) propõem um "modelo de aceitação de bibliotecas digitais" que está ilustrado na figura abaixo:

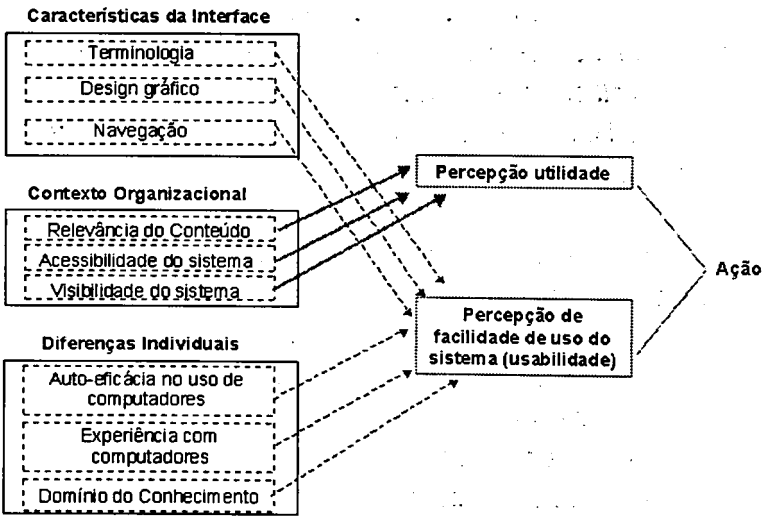


Figura 1 – Modelo de adoção do usuário de bibliotecas digitais (Thong e outros, 2004, 2002).

Este modelo de adoção, composto de nove fatores agrupados em três dimensões, pode ser assim entendido:

1ª. Dimensão – *características da interface* – engloba os seguintes fatores:

• terminologia - as palavras, sentenças, abreviações e linguagem de indexação utilizadas afetam a compreensão dos usuários quanto ao conteúdo e sua estrutura de organização, exigindo ou não um esforço para a atividade de recuperação de informações que pode comprometer a experiência do usuário.

• *design* gráfico - afeta o estilo e a forma como os usuários vão se comunicar com as bibliotecas digitais. Telas organizadas e bem desenhadas ajudam os usuários a escanear e identificar fácil e rapidamente informações relevantes.

• navegação - proporciona contexto e flexibilidade para que o usuário possa encontrar as informações necessárias, independentes da fonte e do local de armazenamento, utilizando dois sub-sistemas – exploração (*browsing*) e busca (*searching*).

2ª. Dimensão – *contexto organizacional* – engloba os fatores:

• relevância do conteúdo - quanto maior for o alinhamento e a adequação do conteúdo às necessidades dos usuários, maior será a percepção deste quanto à utilidade e usabilidade do sistema.

• acessibilidade do sistema - influencia a percepção de facilidade de uso da biblioteca digital e também a sua aceitação.

• visibilidade da biblioteca digital - o grau com que as inovações são comunicadas também podem influenciar a aceitação pelos usuários.

3ª. Dimensão – *diferenças individuais* – engloba os fatores:

• auto-eficácia no uso de computadores - mede o nível de autoconfiança que o usuário tem ao usar uma biblioteca digital não familiar. É o julgamento individual sobre o nível de capacidade de utilizar novos sistemas de informação.

• experiência com computadores – reflete quanto tempo o usuário lida com computadores e seu nível de *expertise* para tal e;

• conhecimento do domínio ou área temática da biblioteca – o co-

nhecimento que o usuário possui sobre o tema tratado pode influenciar sua aceitação das bibliotecas digitais.

Conforme observado no modelo, quando as características da interface propiciam espaço para a valorização das diferenças individuais dos usuários dá-se o aumento da percepção de facilidade de seu uso. Já os fatores incluídos na segunda dimensão – contexto organizacional – exercem impacto diretamente sobre a percepção de utilidade do sistema por parte dos usuários.

Com base nestas discussões, depreendem-se vários e diversificados desafios, características e funcionalidades a serem observadas no *design* de interfaces.

3. Design de interfaces

Arms (2000) ressalta que uma biblioteca digital é tão boa quanto assim for sua interface, pois ela pode facilitar e melhorar a comunicação, reduzir o esforço necessário para compreender a organização estrutural e espacial dos conteúdos, localizar objetos digitais específicos no sistema e nas telas e proporcionar uma navegação fácil. Para Galitz (2002), a melhor interface é aquela que não é notada, que leva os usuários a se focalizarem na informação e na tarefa, sem que percebam o mecanismo usado para apresentar a informação e desenvolver a tarefa.

O *design* de interface é parte integrante do modelo conceitual do sistema[4], juntamente com o *design* funcional (que especifica as funções a serem oferecidas aos usuários) e *design* de metadados (que especifica a estrutura e esquema de organização e descrição do conteúdo) (Arms, 2000).

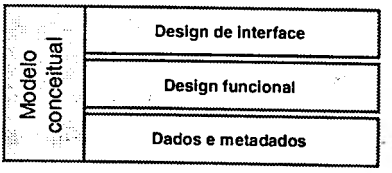


Figura 2 – Modelo conceitual (Arms, 2000)

Deste modo, o projeto da interface que deve ser alinhado com as especificidades, diretrizes e propostas definidas no modelo conceitual é

onde se determina a apresentação dos conteúdos nas telas e a forma de interação a ser oferecida aos usuários para sua manipulação.

Características da interface

Há uma série de características desejáveis em uma interface para bibliotecas digitais, algumas delas (adaptadas de Tedd, Large, 2005) são:

- clareza – falta de ambigüidade em termos visual, conceitual e linguístico;
- compatibilidade – com o usuário final, com a tarefa e também com qualquer versão mais recente do sistema, ou outro tipo similar de sistema;
- consistência – interna em todos os subsistemas da biblioteca em termos das categorias e do vocabulário utilizado. Isto é, as mesmas ações devem sempre dar o mesmo resultado. Assim como consistência externa na entrega da informação;
- flexibilidade – a interface deve prover muitas maneiras do usuário efetuar uma dada operação;
- passividade – a interface deve assumir um papel passivo, permitindo que o usuário detenha o controle da interação;
- visibilidade – os efeitos de ações e objetos devem ser visíveis;
- eficiência – a interface deve minimizar o esforço gasto para executar uma tarefa, por exemplo: reduzir movimentos dos olhos e mãos;
- familiaridade para os usuários, usando conceitos, linguagem e metáforas do mundo real dos usuários;
- diversidade e adaptabilidade – a interface deve suportar convenientemente todas as classes de usuário e ser capaz de identificar individualmente cada usuário e se adaptar a ele, desenvolvendo, interativamente, uma linguagem adequada à interação entre ela e o usuário em questão;
- complacência – a interface deve permitir que o usuário recupere-se de situações de erro, bem como considerar que ele pode esquecer informações já apresentadas; perdoando e prevenindo erros humanos inevitáveis e comuns e provendo mensagens construtivas nestes casos;
- conveniência – a interface deve permitir fácil acesso a todas as operações;

• prestimosidade – a interface deve ser prestativa, fornecendo ajuda quando requisitada ou quando perceber que o usuário se encontra em dificuldades. A ajuda pode se dar na forma de mensagens de erro, conselhos etc., e deve ser clara e precisa, não conduzindo o usuário a situações embaraçosas, nem exigindo dele conhecimentos que ele não tenha ou não possa obter pelo próprio sistema;
• imitação – a interface deve imitar o diálogo humano. Isto não significa necessariamente o uso de linguagem natural, mas sim a exploração de aspectos da comunicação humana não orientados a comandos, tais como o uso de exemplos, explanações, analogias, comparações, descrições etc;
• naturalidade – a interface deve se comunicar com o usuário de maneira natural, não exigindo o conhecimento de terminologia não referente à tarefa;
• satisfação – a interface deve satisfazer o usuário e não gerar frustração. Ela não deve demorar na resposta e deve permitir que o usuário obtenha ajuda em qualquer ponto da interação;
• personalização – a interface deve possibilitar ao usuário interagir com o sistema no sentido de customizá-lo de acordo com seu interesse, perfil ou preferências;
• naturalidade e intuitividade.

Funcionalidades da interface

A interface de uma biblioteca digital pode ter as seguintes funcionalidades (adaptadas de Tedd; Large, 2005, Dreher e outros, 2004):

• promover uma visualização global do conteúdo da biblioteca – a interface deve possibilitar ao usuário a compreensão rápida e geral do conteúdo da biblioteca digital;
• apresentar a informação em múltiplos formatos – desenvolver técnicas e sistemas que suportem diferentes formas de visualização do conteúdo, de acordo com sua natureza e com a necessidade e contexto do usuário (imagens, textos, coleções de textos, técnicas de visualização em duas ou três dimensões, gráficos, diagramas, imagens interativas, realidade virtual, simulações, GIS, vetores espaciais, busca probabilística ou conceitual, mapas conceituais, dentre outras).

• oferecer mecanismos de recuperação de informação – existem duas maneiras para se prover acesso ao conteúdo das bibliotecas, o mecanismo de busca (*searching*) e o de exploração (*browsing*) que significa a navegação por menus de vários tipos, texto e imagem e *menus pull-down*. A apresentação da informação resultante da busca deve ser apresentada de forma clara, consistente e segundo algum critério definido de priorização do conteúdo, habilitando os usuários a selecionarem o que é relevante.

• auxiliar / orientar – interfaces de ajuda tem uma função específica no que tange a facilitar a compreensão dos usuários sobre o sistema, dirimir dúvidas operacionais ou de conteúdos e auxiliar os usuários. Isto pode ser feito de várias maneiras: tour guiados, manuais, orientações etc.

4. Embasamento do design de interface: entendimento dos usuários

Para atingir todo o potencial das bibliotecas digitais e aumentar a probabilidade de sua adoção e uso, o *design* e/ou as melhorias de suas interfaces exigem uma efetiva identificação, entendimento e tradução das necessidades e comportamentos informacionais de seus usuários. Este entendimento pressupõe que o *design* incorpore, dentre suas atividades, de forma contínua e sistematizada, estudos junto aos usuários visando ter melhor compreensão de seus modelos mentais cognitivos em situação de resolução de problemas e tomada de decisão.

Estudar usuários, de acordo com a ciência da informação, pode ser visto sob dois prismas distintos:

• estudos de uso – que se focam nos aspectos de usabilidade do sistema;

• estudos de usuários – que se focam na análise do comportamento interno e/ou exteriorizado pelos sentidos, que permite aos indivíduos construir e projetar seu movimento através do tempo e do espaço. Englobam análises e observações da interação e uso de informação pela comunidade independentemente do sistema.

Estas duas perspectivas de estudos, identificados respectivamente como "centrados no sistema" e "centrados no usuário", são evidenciadas por Savolainen (1993), o qual relata que o pesquisador pode escolher entre dois tipos ideais de pontos de vista ou perspectiva a partir da qual irá entender os problemas de busca e uso da informação. O ponto de vista do intermediário (sistema) ou do usuário (indivíduo) da informação. Cada uma dessas perspectivas implica em diferentes interesses em relação aos processos de busca e uso da informação, pois refletem posições opostas dos atores dentro da cadeia de comunicação (Figura 3).

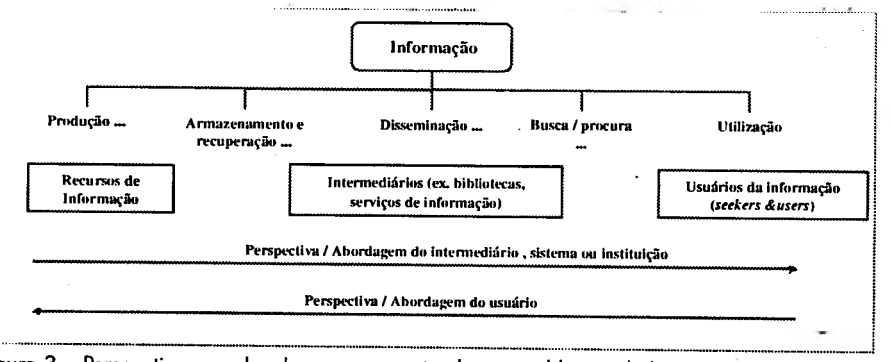


Figura 3 – Perspectivas ou abordagens para entender os problemas de busca e uso da informação (Savolainen, 1993)

Dervin e Nilan (1986) mencionam que o estudo centrado no sistema (também denominado estudo centrado na instituição ou estudo tradicional) foi dominante durante muito tempo. Estudos desta natureza caracterizam todas as funções de comunicação (ex.: armazenamento, recuperação, distribuição, busca e uso da informação) em termos do sistema. Seus questionamentos são focados no uso que os indivíduos fazem do sistema.

Já a perspectiva "centrada no usuário" (ou estudos alternativos) surge a partir da década de 1970, quando se evidencia a premência em ampliar os focos das pesquisas, concentrando-se nos atores individuais dos processos de busca e uso da informação, dentro de contextos sociais, práticos e culturais.

Enquanto os estudos orientados a sistemas examinam o que acontece no ambiente informacional externo ao indivíduo, os estudos orientados a usuários examinam também as necessidades e preferências psico-

lógicas e cognitivas internas do indivíduo e como elas afetam os padrões de procura e comunicação da informação (Choo, 1998).

O desenvolvimento de estudos de usuários pode ser feito com uso de três abordagens:

- Abordagem de Valor Agregado de Robert Taylor
- Abordagem do Processo Construtivista de Carol Kuhlthau
- Abordagem de *Sense-Making* de Brenda Dervin.

Cada uma destas abordagens tem seu foco em fases diferentes do processo de estudos de usuários: a de Taylor se concentra mais na tradução e operacionalização das necessidades informacionais dos usuários para o sistema, enquanto Dervin enfatiza mais a fase anterior a estas atividades, que é a identificação e entendimento de necessidades e comportamento informacional. Já Khulthau enfatiza o estudo e análise do processo de busca e uso da informação na situação ou tarefa específica do usuário referente à pesquisa bibliográfica.

Abordagem do Valor Agregado – Robert Taylor

Esta abordagem considera como "valor agregado" os processos, as características ou os atributos que, uma vez adicionados aos dados e informações, os tornam de maior valor para o usuário.

Desenvolvido na década de 70, este modelo é composto de três módulos: – o usuário ou cliente com seus problemas, necessidades e interesses; – a interface entre o sistema e o cliente, onde fica o denominado "espaço de negociação"[5] e – o sistema propriamente dito, que é definido como um conjunto coerente de processos e atividades que agregam valor à informação e produzem diversos resultados (Taylor, 1986). Sua base de sustentação emerge do que Taylor denomina de "ambiente de uso da informação" (Figura 4).

Este "ambiente de uso da informação" é definido como sendo – um conjunto de elementos que afetam o fluxo e uso da informação para dentro e para fora de entidades definidas ou de grupo de clientes; – e que determinam os critérios através dos quais o valor da informação será julgado. Implica em se descrever a organização, as pessoas e os problemas. (Taylor, 1986). Portanto, o entendimento das peculiaridades, seme-

lhanças e diferenças na busca e uso da informação de grupos específicos será o *input* para a operacionalização de processos executáveis no sistema e na interface que, posteriormente, serão concretizados e se tornarão disponíveis aos usuários na resposta do sistema (*output*) .

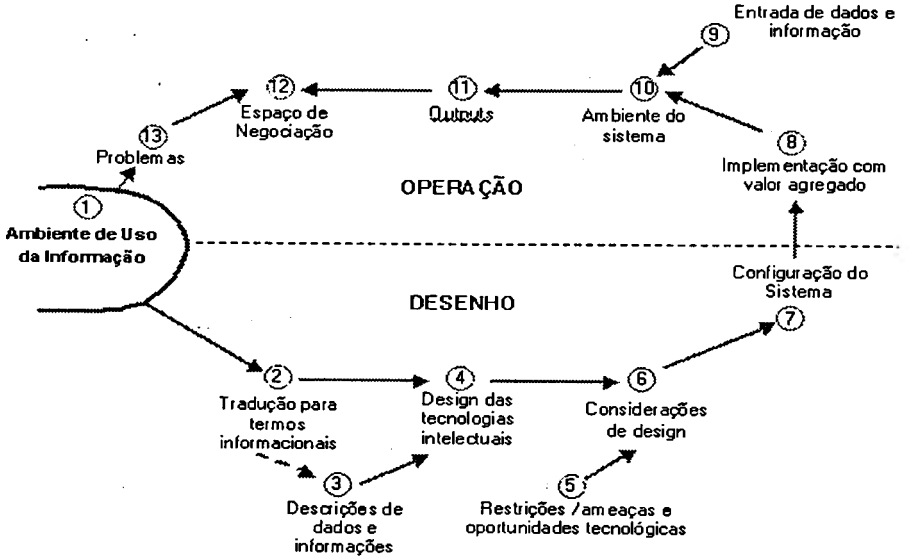


Figura 4 – Modelo de Valor Agregado de Robert Taylor (1986).

Abordagem do Processo Construtivista – Carol Kuhlthau

Esta abordagem enfatiza a ocorrência de estados afetivos e cognitivos que naturalmente surgem num processo de busca da informação. O eixo central da abordagem de Kuhlthau (1991) é o "Processo de Busca de Informações" (*Information Search Process* – ISP) considerado como a atividade construtivista do usuário para encontrar significado a partir da informação, a fim de ampliar seus conhecimentos sobre determinado problema ou assunto. Esse processo ocorre em fases experimentadas pelos indivíduos à medida que constroem sua visão de mundo, através da assimilação de informações novas. A análise destas fases deve incorporar três domínios de atividades: físico (ações realmente executadas pelos usuários); afetivo (sentimentos experimentados) e cognitivo (idéias relacionadas tanto ao processo quanto ao conteúdo).

Entendendo que busca de informação é um processo de elaboração que envolve o conjunto de experiências pessoas e sentimentos, bem como idéias e ações, Kuhlthau (1991) delineia seu modelo identificando as fa-

ses acima mencionadas e detalhando as experiências dos usuários em cada uma delas.

- Fase 1 – Iniciação – fase em que o indivíduo se conscientiza da falta de conhecimento, informação ou entendimento para resolver um problema, enfrentar um desafio, acompanhar um projeto, alocar um recurso ou tomar uma decisão. Os sentimentos que afloram nesta fase são a incerteza, ansiedade e apreensão. Os pensamentos (cognitivo) se referem à contemplação do problema, compreensão da tarefa e relacionamento do problema a experiências anteriores. As ações executadas englobam a discussão e a tarefa mais apropriada é reconhecer a necessidade de informação.
- Fase 2 – Seleção – ocorre quando a tarefa é identificar e selecionar a área ou tópico genérico a ser investigado ou a abordagem a ser feita. Quando a seleção se concretiza e o usuário se sente pronto para iniciar a busca, sentimentos de alívio, otimismo e prontidão para começar a pesquisa são comuns. Cognitivamente ocorre o julgamento da informação frente aos interesses pessoais, requisitos de trabalho e tempo alocado. A ação tomada é a de conferir a informação com outras pessoas. A tarefa mais apropriada é a de identificar um tópico genérico.
- Fase 3 – Exploração – fase mais difícil para os usuários e de maior complexidade para os fornecedores e *designers* de sistemas e serviços de informação. A tarefa aqui é investigar informação dentro do problema geral, a fim de aumentar o entendimento pessoal sobre o mesmo. A inabilidade de expressar precisa e claramente a informação que se necessita impacta na comunicação entre o usuário e o sistema ou mediador. Os sentimentos comuns são confusão, incerteza, dúvida, inadequação e frustração com o sistema, principalmente quando a informação é inconsistente ou incompatível. Pode ocorrer vontade de abandonar todo o processo. Em termos cognitivos, o usuário busca tornar-se suficientemente informado e orientado sobre o tema, para formar o foco ou o ponto de vista pessoal. A ação mais comum é localizar a informação sobre o tópico em geral e estender o entendimento pessoal sobre o assunto.
- Fase 4 – Formulação - fase em que ocorre o *turning point* (ponto de virada) do processo de busca da informação. Os sentimentos de in-

certeza diminuem à medida que o entendimento aumenta e surge uma sensação de maior clareza. A tarefa aqui é formar um foco a partir da informação encontrada na fase anterior do processo. A ação mais comum é ler as anotações sobre os temas e a tarefa mais apropriada é formular foco a partir da informação encontrada.

• Fase 5 – Coleta - fase onde a interação entre usuário e sistema de recuperação de informação se torna mais efetiva e eficiente. É a coleta de informações para o problema em foco. A sensação é de uma direção mais clara, proporcionando maior confiança para o aprofundamento do tema. O interesse aumenta e pensa-se em definir, estender e dar apoio ao tópico selecionado. Fisicamente, a ação mais comum é a de procurar informação relevante e focada e a tarefa mais apropriada é reunir informação pertinente ao foco.

• Fase 6 – Apresentação - a tarefa é completar a busca e resolver o problema em questão. Os sentimentos são de alívio e satisfação. O estado cognitivo está mais claro e focado. A ação mais comum é sintetizar de forma personalizada o tópico ou o problema e a tarefa apropriada refere-se a completar a busca e preparar-se para apresentar ou usar seus resultados.

Abordagem Sense-Making - Brenda Dervin

A abordagem "*Sense-Making*"[6] busca avaliar como os usuários percebem, compreendem e sentem suas interações cotidianas e como usam a informação e outros recursos neste processo. Tem como foco de estudo o processo de construção de sentido, definido por Dervin (1983) como a atividade humana de observação, interpretação e compreensão do mundo exterior, inferindo-lhe sentidos lógicos, advindos do uso de suas experiências.

Esta abordagem enfatiza o aspecto processual de construção da informação pelo indivíduo, e considera que, na busca de criação de significado, as pessoas se movem ao longo de um *continuum* de tempo-espaço em constante mutação.

Este modelo se operacionaliza a partir de três instrumentos especificamente desenvolvidos:

• Instrumento 1 – metáfora da construção de sentido ou trinômio do *Sense Making* – composta pelos elementos - *Situação:* contexto no

tempo e no espaço onde o sentido é construído e a necessidade de informação emerge. *Gaps ou Lacunas*[7] - pontos em que o usuário tem pouca ou nenhuma compreensão do processo e isto o leva a interromper seu caminho. *Usos/help* – se refere ao uso ou não uso da informação pelo indivíduo na busca de sentido. Operacionaliza-se na forma como as respostas, que o sistema provê, atendem (ou não) às necessidades de informação dos usuários. Esta metáfora pode ser assim representada:

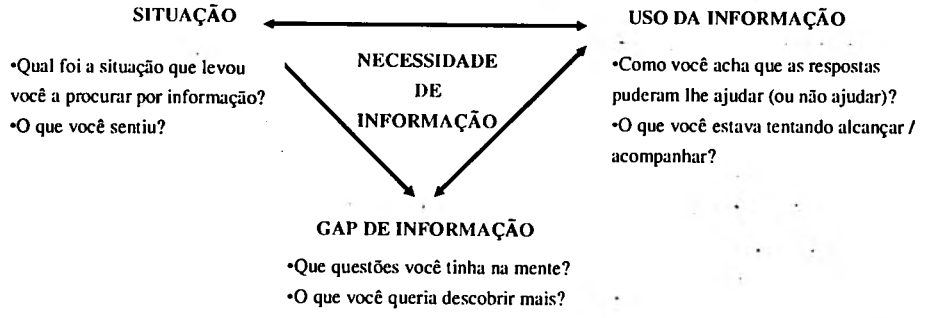


Figura 5 – Metáfora da Construção de Sentido ou trinômio do Sense-Making.

• Instrumento 2 – Operacionalização das variáveis do Trinômio – como consolidação dos estudos desenvolvidos por Dervin e seguidores durante os últimos 13 anos, já foi possível a construção de uma série de categorias para codificação e análise das situações, lacunas e usos passíveis de serem aplicados em estudos de qualquer natureza.

• Instrumento 3 – métodos de coleta de dados (5 métodos) – esta abordagem, dada a sua especificidade na forma de coletar dados, desenvolveu até o momento cinco técnicas de entrevistas distintas para atender a diferentes propósitos: (1) *Micro-Moment Time-Line Interview*; (2) *Neutral Questioning*; (3) *Helps/Hurts Chaining*; (4) *Close-ended Sense-Making Interview*; e (5) *Message Q/ing Interview*. As principais e mais utilizadas técnicas são as duas primeiras.

Estudos de uso/usabilidade

Os estudos de uso, denominação dada por Dervin e Nilan (1986) para os estudos de usuários tradicionais, ganham espaço e notoriedade

quando o conceito de usabilidade passa a ser foco de pesquisa também de várias áreas do conhecimento como a ciência da computação, psicologia, ergonomia dentre outras.

As perspectivas sobre usabilidade têm evoluído ao longo do tempo, passando de uma preocupação mais centrada no desenvolvimento de artefatos computacionais para atendimento a necessidades físicas dos usuários e, atualmente, aparece mais centrada às necessidades e capacidades cognitivas dos usuários durante o processo de interação com os sistemas (Arms, 2000).

De acordo com a ISO 9241-11 (*Guidance on usability*), usabilidade é a "extensão que um produto pode ser usado por usuários específicos, para alcançar objetivos determinados de efetividade, eficiência e satisfação em um contexto de uso definido".

Ainda não existe consenso para a determinação de critérios de usabilidades especificamente aplicados para bibliotecas virtuais, muito embora Saracevic (2004) mencione que tais estudos têm sido largamente utilizados para suas avaliações.

Dentre os vários estudos já desenvolvidos (Dias, 2003; Blandford; Buchanan, 2003; Tedd; Large, 2005), os critérios mais utilizados (muitos deles definidos com base nos critérios de Jacok Nielsen, 1994) são:

- conteúdo: acessibilidade, disponibilidade, claridade, complexidade, grau em que é informativo, transparência, entendimento e esforço de entendimento, adequação, qualidade, validade, dentre outros.
- processos: que se refere a conseguir realizar busca, exploração (*browsing*), avaliar e obter um recursos. Os critérios envolvem a capacidade para aprender e o esforço para utilizar, conveniência e facilidade de uso, apoio para executar ações, dificuldade de interpretação, taxa de erros, dentre outros.
- formato: atratividade, esforços de sustentação, consistência, representação de conceitos, dentre outros.
- avaliação geral: satisfação, sucesso, qualidade da experiência, barreiras/irritabilidade, dentre outros.
- funcionalidades do sistema: recursos, produtos e serviços disponíveis para a execução de distintas tarefas.
- uso: utilzação do sistema em relação as tarefas, locais, horários e outros.

- habilidade de aprendizagem *(learnability)*: medida importante para que as interfaces de bibliotecas digitais possam melhor apoiar os processos mais complexos de acesso, busca e uso da informação.

Estudos de usabilidade em bibliotecas digitais podem ser realizados utilizando técnicas variadas, tais como: levantamentos diretos e *on-line*, entrevistas estruturadas, *focus group*, observações, acompanhamento de tarefas, estudos de casos, análises de *log*, análise de registros e de uso da biblioteca digital, documentos, análise da comunicação (antropologia). Cada uma destas técnicas tem suas aplicações, forças e fraquezas.

5. Considerações finais

O design de interfaces de bibliotecas digitais sofrerá as influências e impactos da evolução de conceitos e abordagens tais como a computação ubíqua, a mobilidade, acessibilidade e as estratégias de integração; e também o amadurecimento dos usuários na busca e uso de informação digital.

Neste contexto, é importante lembrar que, apesar de termos dado foco neste capítulo, a interface para usuários finais, entendemos ser essencial a análise e consideração das mesmas questões para as interfaces necessárias a todos os atores do processo de uso das bibliotecas digitais (criação, publicação e comunicação científica), tais como produtores, gestores e distribuidores de conteúdo e de fontes.

Todas as tendências caminham para ampliar, qualitativamente, o acesso, o uso e o compartilhamento da informação entre usuários e comunidades, possibilitando a interação com a informação a partir de qualquer dispositivo, hora e local. Um dos grandes desafios para a interface, neste contexto, é a necessidade de auxiliar os usuários a lidarem com a multiplicidade de formatos e suportes das informações em contraposição com a sobrecarga cognitiva imposta por tarefas cada vez mais complexas e interdisciplinares.

Notas

[1] O termo Interface Gráfica com o Usuário (Graphical User Interface – GUI) é usado para descrever uma interface homem/máquina quando imagens gráficas e analogias de gestos humanos formam a linguagem básica de interação entre o usuário e o computador.

[2] Inteligência Artificial: uma definição bastante esclarecedora é a de E. Rich (1991) para a qual, Inteligência Artificial (IA) é o "estudo de como fazer os computadores realizarem tarefas as quais, até o momento, os homens fazem melhor". Ou ainda a definição de J. L. Laurière (1990): "Todo problema para o qual nenhuma solução algorítmica é conhecida, é um problema da IA". Ou seja, as tarefas relacionadas com o processamento simbólico, reconhecimento de imagens e tudo o que envolva "aprendizado". (http://www.universiabrasil.net/pesquisa_bibliotecas/materia.jsp?id=6310\)

[3] Adaptatividade: refere-se à habilidade do sistema se adaptar a um ambiente mutável, a diversos grupos de usuários e diferentes padrões de uso. Essas adaptações podem se dar no conteúdo ou na interface de usuário. Stephanidis e outros (2000).

[4] Modelo conceitual significa a descrição do sistema proposto em termos de uma série de idéias integradas e conceitos sobre o que ele deve fazer, se comportar e parecer de forma a ser entendido pelos usuários na maneira em que eles pretendem e planejam.

[5] Espaço de Negociação: é onde o "casamento" entre sistema e usuário ocorre e é avaliado. Este espaço reflete as preocupações ocorridas na análise do ambiente de uso da informação, que se denominam "problemas" (Taylor, 1986).

[6] Segundo Ferreira (1997), existe a convenção de se utilizar o termo *Sense-Making* com as iniciais maiúsculas e hífen para se referir à abordagem e *sense making* em minúsculas para se referir ao fenômeno.

[7] As lacunas, segundo Ferreira (1995) podem ser entendidas como: (1) um estado anômalo do conhecimento (Belkin; Oddy; Brooks, 1982); (2) um estado de incerteza (Krikelas, 1983); ou uma situação na qual o indivíduo está tentando chegar à compreensão de alguma coisa (Dervin, 1983).

Referências

ARMS, William Y. *Digital Libraries*. Massachussetts Institute of Technology, 2001.

BEAUDOUIN, Michel Lafon. Designing Interaction, not Interfaces. *Coomunication of the ACM*, AVI, maio, 2004.

BELKIN, N. J.; ODDY, R. N.; BROOKS, H. M. ASK for Information Retrieval:

Part I - Background and theory. *Journal of Documentation*, v. 38, n.-2, jun, 1982. p. 61-71.

BLANDFORD Ann; BUCHANAN, George. *Usability of digital libraries*: a source of creative tensions with technical developments. TCDL Bulletin, 2003. Disponível online: http://www.ieeetcdl.org/Bulletin/current/blandford/blandford.html

BLANDFORD, Ann; BUCHANAN, George. Usability of digital libraries: a source of creative tensions with technical developments. UCL Interaction Centre Interaction Design Centre, 2003.. Disponível online em http://www.ieee-tcdl.org/Bulletin/current/blandford/blandford.html . Acessado em Abril, 2005.

CHOO, Chun Wei. *The knowing organization:* how organizations use information to construct meaning, create knowledge and make decisions. Oxford University Press, (1998).

DERVIN, Brenda. *An overview of Sense-Making research:* concepts, methods, and results to date. Paper apresentado no International Communication Association Annual Meeting, Dallas, texas, Maio, 1983. Disponível em http://communication.sbs.ohio-state.edu/sense-making/art/artdervin83.html . Acessado em Set. 2002.

DERVIN, B.; NILAN, M. Information needs and uses. Annual Review of Information Science and Technology (ARIST), v. 21, 1986.

DIAS, Cláudia. *Usabilidade na Web*: criando portais mais acessíveis. São Paulo: Alta Books, 2003.

DILLON. *Technologies of information:* HCI and the digital library. In: J.M. Carroll, ed. Human-Computer Interaction in the New Millenium. ACM Press, 2002.

DREHER, Heinz; KROTTMAIER, Harald; MAURER, Hermann. What we expect from Digital Libraries. *Journal of Universal Computer Science*, v. 10, n. 9, 2004.

GALITZ, W.O. *The essential guide to user interface design*. 2. ed, Ed. Wiley, 2002.

HORGEN, Svend Andreas. *Essay on intelligent user interfaces and adaptivity*, November, 2001.

KRIKELAS, J. Information-seeking behavior: patterns and concepts. *Drexel Library Quarterly*, v. 19, n. 2, 1983. p. 5-20.

KUHLTHAU, Carol C. Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, v. 42, n. 5, p 361-371, 1991.

MORAN, T. The Command Language Grammars: a representation for the

user interface of interactive computer systems. *International Journal of Man-Machine Studies*, v. 15, 1981. p. 3-50.

NIELSEN, J. *Usability Engineering*. San Francisco, USA: Morgan Kaufmann Publishers, 1994.

PRESSMAN, R. S. Software Engineering: A Practioner's Approach. 3. Ed., McGraw-Hill, 1992.

SAVOLAINEN, Reijo. The Sense-Making Theory: reviewing the interests of a user-centered approach to information seeking and use. *Information Processing and Management*, v. 29, n. 1, 1993. p. 13-28.

STEPHANIDIS, C; AKOUMIANAKIS, D.; PARAMYTHIS A.; NIKOLAOU C. *User interaction in digital libraries:* coping with diversity through adaptation - International Journal of Digital Libraries, 2000.

TAYLOR, Robert S. *Value-added processes in Information Systems*. Ablex Publishing Corporations, 1986.

TEDD, A. Lucy.; LARGE, Andrew. *Digital Libraries:* Principles and Practice in a global environment. K. G. München, 2005.

THONG, James Y.L.; HONG, Weiyin; TAM, Kar Yan. What leads to user acceptance of digital libraries? *Communications of the ACM*, v. 47, n. 11, Nov, 2004.

THONG, James Y.L.; HONG, Weiyin; TAM, Kar Yan. *Understanding user acceptance of digital libraries:* what are the roles of interface characteristics, organizational context, and individual differences? International Journal of Human-Computer Studies, 2002. Disponível online http://www.idealibrary.com.on. Acessado em Fev. 2005.

TSAKONAS, G, KAPIDAKIS, S.; PAPATHEODOROU, C. A new interaction evaluation framework for digital libraries. Paper apresentado no DELOS Workshop on the Evaluation of Digital Libraries, 2004.

Tornando a informação disponível: o acesso expandido e a reinvenção da biblioteca

"A biblioteca é um conceito, tanto quanto um lugar – é função, não apenas forma"[1]

*Marília Levacov**

1. Introdução

A transição entre a Revolução da Imprensa, iniciada por Gutenberg (que não acabou ainda) e a Digital (que está apenas iniciando), coloca o profissional da informação entre dois modos eventualmente conflitantes de encarar sua profissão, o local onde a exerce e as atividades e responsabilidades que a acompanham[2].

De um lado, uma abordagem que lida com a informação que existe, principalmente como tinta sobre o papel, que a define e delimita, facilitando sua a classificação e armazenamento, pois para elas existem critérios e convenções muito bem definidos. De outro lado, a mudança progressiva e

* Doutora em tecnologia e mídia pela Boston University (Boston, 1994) e mestre em computação na educação pelo Lesley College Graduate School (Cambridge, 1986). Professora e pesquisadora na UFRGS até 2003, onde também coordenou o Centro de Informação em Ciência e Tecnologia do Instituto Latino-Americano de Estudos Avançados. Trabalha atualmente em consultorias privadas. mlevacov@penta.ufrgs.br

inevitável desta informação para a tela como arquivos digitais, alterando os sólidos paradigmas informacionais construídos através de séculos.

O espaço da biblioteca e a formação do próprio profissional da informação como o conhecemos organizaram-se e evoluiram em função das características físicas da informação em átomos. A biblioteca era e ainda é, para muitos, um local onde reside um objeto analógico (o livro, o documento etc.) e para onde se dirigem os usuários em busca de informação. Prateleiras para armazenamento, critérios de indexação, catálogos, normas técnicas, serviços de empréstimo, subdivisões de funções etc. criaram-se em virtude destas características físicas.

A crescente desmaterialização da informação, em decorrência da mudança do suporte analógico para o digital, e as conseqüências desta mudança (a explosão de informação e a tela como suporte, por exemplo) requerem do profissional desta área novos conceitos sobre o que é e faz, onde reside o objeto de seu trabalho, quem o possui e também sobre quem são e onde estão os usuários desta informação que *"torna-se cada vez menos ligada ao objeto físico que a contém"*[3].

À medida que o mundo se move, mais e mais, em direção à digitalização (impulsionado pela era de avanços tecnológicos em que vivemos), as instituições que se relacionam com a informação, tais como as bibliotecas, precisam transformar-se também para garantir sua acessabilidade e manter-se em dia com as necessidades informacionais da sociedade a que atendem. Caso não o façam, correm o risco de se tornarem irrelevantes para grande percentagem de usuários, uma vez que, a maioria das informações atualizadas, principalmente em determinadas áreas do conhecimento, nasce e permanece sob a forma digital, sem nunca alcançar uma versão impressa. Portanto, é sob esta forma digital, instável e volátil, que ampla e significativa parte do discurso humano se apresenta, cada vez mais, no início do século XXI[4].

Isso certamente não significa que a biblioteca "tradicional" vá desaparecer num futuro próximo em virtude do crescimento da importância da informação digital, mas que inegavelmente ela tende a se "desmaterializar" e "descentralizar", evoluindo junto com a sociedade onde se insere.

Essa alteração no conceito tradicional de biblioteca, portanto, não é uma ocorrência isolada, mas o resultado das transformações que a sociedade

globalizada e interconectada está vivendo e que se expressam de diferentes maneiras, mas com grande pungência nas atividades relacionadas com a informação.

> Não devemos ignorar os problemas políticos e culturais decorrentes desta informatização que, a longo termo, possuem importantes conseqüências para bibliotecas, bibliotecários e usuários.[5]

A sociedade pós-industrial e suas necessidades especiais foram, portanto, responsáveis pela criação do que parece ser uma dicotomia entre tradição *versus* novas tecnologias. Esta dicotomia é alavancada pelo desenvolvimento e barateamento da tecnologia digital (computadores, memória para armazenamento, redes de comunicação, *softwares* intuitivos e a progressiva ampliação da largura de banda disponível), aliados a explosões de publicações impressas e digitais, certas características da informação digitais (como sua facilidade de duplicação), paralelamente a fatos como o encarecimento do papel (e as questões associadas com sua sustentabilidade), com a degradação física de coleções construídas sobre suportes tradicionais e ainda das dificuldades financeiras e físicas de administrar crescentes volumes de material impresso[6].

O formato digital permite que a informação resida no que chamamos de bibliotecas digitais ou virtuais, mas desafia o profissional da área a deixar de pensar como um "guardião da informação", entendendo as caraterísticas especiais destas coleções e desenvolvendo estratégias informacionais que tirem proveito destas tecnologias ao mesmo tempo em que repensam certos aspectos de sua profissão. Os aspectos referentes ao uso das novas tecnologias, principalmente, tornam-se elementos importantes e precisam fazer parte do conhecimento deste profissional e do planejamento de uma biblioteca virtual, não mais uma atividade secundária a ser relegada aos técnicos dos CPDs. A compreensão destes aspectos permite ao profissional da informação tomar as decisões adequadas e desenvolver as políticas necessárias à correta preservação do acervo e de seu acesso.

Muitos são os tópicos importantes quando se reflete sobre a emergência das bibliotecas digitais. A maioria destes será tratada individualmente nos demais capítulos. Neste, faremos breve revisão daqueles itens que nos parecem mais pontuais.

2. A biblioteca híbrida e a questão da preservação (da informação e do acesso)

O conceito para acomodar estes dois momentos paralelos que estamos vivendo é o da "biblioteca híbrida", aquela que contém não apenas material impresso e magnético, mas também informação digital (em múltiplos formatos: mídias óticas *on* e *off-line*, como CD-ROMs e DVDs, terminais para acesso a catálogos, a bancos de dados e a alguns dos variados tipos de documentos digitais). Mais e mais, a maioria das bibliotecas precisa incluir, voluntariamente ou não, estes tipos de documentos em suas coleções. Tais bibliotecas podem possuir também computadores para acesso de seus usuários à Internet. Os profissionais a que lá atendem necessitam, no mínimo, adquirir algum conhecimento sobre a operação dos computadores e desenvolver estratégias para instrução destes conhecimentos a seus usuários.

É importante lembrar também que o fato de a maioria das bibliotecas possuir ainda informação primordialmente em átomos não as isenta de terem de lidar com os problemas referentes ao formato digital. Mesmo que por razões financeiras ou outras quaisquer não adotem catálogos eletrônicos não possuam documentos computadorizados em seus acervos nem forneçam acesso a bases de dados digitais, terão de (além daquele de perceber os usuários cada vez menos interessados nos documentos que abrigam) enfrentar um problema agudo relativo às publicações impressas ao longo do século XX: a preservação de suas coleções.

A maioria dos documentos impressos que residem atualmente nas bibliotecas foi feita com material altamente perecível, a polpa de árvores amaciada e alvejada com ácido clorídrico, que, em poucas décadas (anos nos casos de jornais), progressivamente degrada o papel, até que este se torne um farelo irrecuperável. Grandes esforços estão sendo feitos para desacidificar coleções, mas, até agora, as solução são extremamente caras. Os livros impressos antes da criação e disseminação deste processo industrial, isto é, antes do final do século XIX, não sofrem essa ameaça, mas raras são as bibliotecas que dispõem desses volumes e, mesmo a esses, a versão digital facilita o acesso de usuários remotos, bem como sua preservação.

O uso de papel não-acidificado, apesar de este processo não encarecer o preço final, ainda não se tornou hábito pelas editoras. Na maioria das coleções atuais, não são os livros de grandes autores clássicos que estão em perigo maior (pois estes são reeditados periodicamente e/ou

podem ser readquiridos, se houver disponibilidade financeira), mas aqueles outros, que, de alguma forma especial ou raros, poucas chances terão de reimpressão: documentos únicos, seja pela informação que contém, seja pela impressão aquela em particular, seja pelo tipo de diagramação, ilustração, encadernação, que precisam ser preservados de forma especial, mas também precisam estar disponíveis para o acesso dos usuários, o que nem sempre é compatível com os cuidados que inspiram.

Preservar a informação existente nestas coleções e situações, processo antes feito de modo pouco satisfatório em microfilmes, ganha alento com o barateamento e a popularização das novas tecnologias, com seus equipamentos digilitadores, a flexibilidade e o pouco custo de armazenamento da informação digital, a evolução das interfaces icônicas que não requerem mais conhecimentos matemáticos específicos para o manuseio das máquinas, a sociedade em redes, o crescimento dos usuários remotos, etc., fatos estes reforçados, como mencionamos na acima, pelo custo crescente do livro impresso, pela dificuldade em expandir fisicamente as áreas de armazenamento, por políticas editoriais discutíveis e, principalmente, porque o suporte impresso engloba agora apenas parte da informação produzida atualmente, parte essa menos atualizada e em formato estático, em contraposição aos formatos dinâmicos das hipermídias.

Estes e outros fatos conectados implicaram uma das mais importantes decisões a considerar quando orçamentos de bibliotecas são planejados: aumentar a coleção ou o acesso? A resposta a essa pergunta rapidamente está se tornando clara[7].

3. A descorporificação da informação, da biblioteca e do usuário

A informação sempre existiu em muitos formatos mesmo antes de se tornar digital. Cada um destes apresentava vantagens e desvantagens como portabilidade, durabilidade, confiabilidade, acessabilidade, legibilidade, potencial ou disponibilidade para armazenamento, recuperação e distribuição, custo, etc. Isso fez com que diferentes usos e tipos de informação fossem privilegiados ou não, conservados ou não, em função destas e outras características. Não há discussão, porém, de que, nos últimos 500 anos, a maioria do conhecimento humano existiu como documento impresso sobre o papel.

Uma situação igualmente conflitante e ameaçadora para alguns aconteceu quando os primeiros livros impressos surgiram, uma vez que os "profissionais da informação" da época, os escribas nos mosteiros, consideraram essa forma perigosa e inadequada (poderia expalhar idéias não-censuradas, conter e difundir erros ortográficos, tiraria o poder da mão de quem o deteve por muitos séculos etc.). É a célebre frase do personagem Frollo, de Vitor Hugo: *Ceci tuera cela*[8], sem a qual Umberto Eco considera incompleto qualquer texto sobre a transição da informação do suporte analógico para digital. Apesar de desaprovar o livro impresso e tudo o que representava, a Igreja não conseguiu deter sua proliferação e sua crescente importância como suporte da informação na cultura ocidental, a ponto de se tornarem, livro e conhecimento, simbolicamente sinônimos.

A Era da Informação, mais uma vez, reabre o debate territorial sobre vantagens e desvantagens de um formato e sua inexorabilidade. Desta vez, o digital. A informação neste formato apresenta grandes vantagens porque descorporificada, descolada do papel, habita livremente, como sinal elétrico, o ciberespaço: um ecossistema informacional (formado por computadores, cabos telefônicos, fibras óticas, ondas eletromagnéticas, satélites, telefones fixos ou celulares e outras tecnologias afins) que existe (visível e invisível) ao nosso redor e que cresce exponencialmente a cada segundo, em importância e em volume, estando potencialmente disponível a qualquer um em qualquer hora e em qualquer lugar.

A biblioteca, antes "ilha" de informações visitada fisicamente por usuários que buscavam um produto, principalmente o livro, precisa agora integrar-se e disputar espaço num "mar de informações". Seus usuários, tornando-se progressivamente remotos e descorporificados, acostumam-se àquilo que se pode obter "aqui" e "agora": a informação digital "instantânea" disponibilizada pelas redes, em especial pela Internet e pela crescente largura de banda disponível.

4. Características da informação no suporte digital

Convergência de mídias

Como os dados são processados no computador por meio de impulsos elétricos, para serem lá administrados, precisam todos estar representados de forma "digital" ou, mais precisamente, binária. Chama-se binário

ao sinal elétrico que pode ter apenas dois estados: *on* e *off*, com ou sem corrente, também representado por zeros e uns. Exemplo: 00011010.

Mídias antes residentes em suportes variados (papel de polpa de árvores, de linho, de arroz, pergaminhos, papiros, plástico, telas gessadas etc.), como a maioria das mídias estáticas, ou em suporte magnético, como as mídias dinâmicas (áudio, vídeo etc.), podem ser todas representadas pelos mesmos sinais digitais e, portanto, combinadas, duplicadas, arquivadas, recuperadas, cruzadas e distribuídas de modos inéditos.

Esta mídia digital – ao contrário das mídias anteriores, nas quais cada cópia implicava atenuação do sinal que descrevia a informação (fotocópias, duplicação magnética de fitas de áudio e vídeo etc) – pela precisão de seu suporte tecnológico, permite duplicar e recombinar documentos de forma a tornar indistinguível o original da cópia. Vale lembrar que, como diversas coleções ainda continuam (e continuarão por algum tempo) em formato analógico, muitas bibliotecas precisarão adaptar-se e desenvolver as estratégias necessárias para administrar dois acervos semelhantes ou até redundantes: um analógico e outro digital.

Alteração nos conceitos de "lugar" e de "tempo"

Graças às tecnologias das telecomunicações em rede, onde o documento reside, como *bits* salvos em algum HD em um ponto da rede, deixa de ser importante. Sob a forma digital, a informação passa a habitar um espaço "virtual" e ser visualizada na tela do computador. Por esta razão, a biblioteca digitalizada pode apontar para as fontes de informação sem, necessariamente, possuir a propriedade física das mesmas. O importante passa a ser a disponibilização do acesso e, com freqüência, a confiabilidade da informação disponibilizada. Dada a facilidade de manipulação deste tipo de informação, torna-se agora especialmente importante saber quem a produziu, quem a identificou como valiosa, quem a selecionou para disponibilizar e quem garante sua autenticidade.

O conceito de tempo também se altera, criando-se uma "relativização" do mesmo. Uma de suas maiores conseqüências é que a instantaneidade passa a ser a palavra de ordem. Soluções como o Programa de Comutação Bibliográfica, por exemplo, não mais atendem às necessidades de acesso imediato aos documentos que a informação em rede oportuniza. O volume e a diversidade de informação disponível *on-line*, em um número cres-

cente de áreas, permite que o usuário impaciente possa migrar para uma outra, igual ou similar, com um "clique" de seu *mouse*. Essa biblioteca digitalizada pode estar "aberta" 24 horas por dia, sete dias por semana, competindo em acesso com outras bibliotecas e outras fontes, a maioria bastante anárquicas.

A velocidade das mudanças tecnológicas que acontece nestes ambientes, que podem ser medida em meses às vezes, comparada a processos equivalentes nas mídias anteriores que demoravam séculos, cria confusão e *stress* entre bibliotecários e usuários. Não há sequer consenso em prever quais serão os formatos existentes e as necessidades de uma biblioteca daqui a uma década, o que dificulta enormemente os planejamentos na área.

Novos conceitos de catalogação

Este novo lugar, o ciberespaço, abriga uma multitude de entes descorporificados: documentos diversos, informações em diferentes formatos e de diferentes naturezas, além de pessoas utilizando diferentes protocolos de acesso. Esses protocolos privilegiam determinados tipos de informação, evoluindo e transformando-se com desconcertante velocidade. É essa descorporificação que permite que, de qualquer lugar do mundo onde haja o potencial de acesso, em qualquer hora e de diferentes equipamentos, usuários busquem os locais onde reside – naquele momento – a informação digital desejada.

Uma das ferramentas clássicas para a localização da informação na biblioteca tradicional, o catálogo, é fortemente influenciada por tudo isso. Por séculos, os catálogos representaram uma ferramenta de identificação e descrição que atendia adequadamente às coleções relativamente estáticas ou moderadamente dinâmicas contidas numa determinada instituição. Tais catálogos permitiam encontrar um livro por seu título, autor ou gênero, porque apontavam para a localização física desta informação na prateleira da biblioteca. Para o profissional da ciência da informação, a convergência de mídias, oportunizada pelo formato digital (bem como sua localização volátil), apresenta novos desafios também ao exigir que sejam desenvolvidas novas formas de descrever e indexar estes documentos dinâmicos, em múltiplos formatos e em localizações remotas sobre as quais, muitas vezes, tem-se muito pouco controle.

Por sua importância, o catálogo foi a primeira coisa a ser digitalizada em muitas bibliotecas e, posteriormente, a ser disponibilizada na Internet. Chamado em inglês de OPAC – *Online Public-Access Catalog* (Catálogo de Acesso Público em Rede), referia-se aos documentos analógicos presentes em uma determinada instituição, muitas vezes "reunindo" coleções dispersas por diferentes lugares em uma única longa e flexível prateleira virtual. A maioria das bibliotecas universitárias o possui, em seu *website*, junto com as opções de acesso a bases de dados digitais, locais ou remotas.

A complexa questão de criar um catálogo que descreva e aponte para documentos digitais dinâmicos e inconstantes, em hipermídia, em um volume crescente, será abordada no capítulo sobre metadados deste livro. Mas, enquanto eu escrevia este, a conhecida lista de discussão de Aldo Barreto apontou[9] para um memorável artigo, "*The Future of Cataloging* "[10], onde autora questiona-se sobre se há futuro para atividades de catalogação, indexação etc. numa época de indexadores automáticos, que criam instantanemente o que se poderia chamar de uma lista personalizada que aponta diretamente para a informação dentro da topologia específica do ciberespaço. A autora faz as perguntas duras, mas necessárias, ao descrever, de modo realista, o que ela e outros autores percebem quando conversam com usuários da Internet: que estas pessoas utilizam material digital *on-line*, diretamente, sem recorrer aos catálogos convencionais, acessando documentos que, freqüentemente, não existem ou não são sequer solicitados na biblioteca em átomos. A autora também descreve o formidável projeto Google de digitalização de obras em textos integral (existem outros mais antigos, como o Projeto Gutenberg, que foi o primeiro, mas esses não possuem um indexador automático e poderoso como o Google na função também de patrocinador do projeto[11]). Neste universo, o nível intermediário para localização do texto completo, a função da catalogação e indexação, talvez se torne redundante ou desnecessária e precisa ser repensada.

5. Alguns dados para pensar

"Cada mídia traz seus desafios e, no entanto, a sociedade se adapta". Paul Saffo[12]

Captura, armazenamento, distribuição e seu impactos sobre as políticas de propriedade intelectual

Muitas coisas serão ditas sobre isso nos capítulos adiante mas é importante refletir sobre uma em especial: as redes cliente a cliente. Duas das características de um documento digital – sua duplicabilidade sem sofrer atenuação e sua portabilidade – propiciam que estes sejam distribuídos e compartilhados de modos alternativos e sem controle. Atualmente isso se dá, cada vez mais, por meio das diversas redes P2P *(peer to Peer* – parceiro a parceiro, ou cliente a cliente, como são chamadas em português) e que, como o nome mostra, implicam relação de igualdade entre os dois ou mais usuários envolvidos no compartilhamento dos arquivos. A tecnologia P2P é uma maneira eficiente de transferir informação eletrônica, via rede, possuindo, entre outras vantagens, o potencial de reduzir substancialmente o tempo e os custos, quando comparada com os sistemas de distribuição baseados em servidores.[13]

O mais recente destes protocolos a ganhar notoriedade foi o Napster. Neste modelo, os computadores dos usuários cadastrados em um *site* "organizador" faziam um *upload* de sua lista de arquivos disponíveis (neste caso, principalmente trilhas sonoras em formato MP3) para o banco de dados do servidor Napster e, em troca, acessavam o "catálogo" do mesmo banco para descobrir onde residiam arquivos de seu interesse e quais deles estavam *on-line* e disponíveis naquele momento. Recebiam também informações sobre o tipo de conexão que cada um destes potenciais depósitos possuiam (linhas discadas, dedicadas etc), bem como outros dados que lhes permitisse escolher uma fonte de confiança (outro computador com boa conexão) para seu *download*, à qual, depois, conectavam-se diretamente.

Imediatamente, as gravadoras, as maiores interessadas nos direitos autorais desse tipo de documentos, e alguns dos autores (por meio da *Records Industry Association of America* – RIA) entraram com diversas ações legais e conseguiram gradualmente a extinção do *site* organizador. Claro que um novo *site* poderia surgir imediatamente em outro lugar no ciberespaço, mas, a esta altura, novos protocolos P2P surgiram, mais eficientes e totalmente descentralizados, burlando outra vez as tentativas de controlar o trânsito de informações na rede.

O novo modelo P2P, independe de um servidor, é totalmente descentralizado, funcionando com o conceito de "horizonte". Cada máquina com

este tipo de protocolo vasculha a Internet em busca de outros usuários do mesmo programa (ou de compatíveis), que também estejam sinalizando sua disponibilidade, até um limite de quatro mil usuários, por exemplo. Estes quatro mil estão "enxergando", por sua vez, um outro número semelhante de máquinas (bem como os arquivos que elas estão querendo disponibilizar) e assim sucessivamente. O usuário então dispara sua(s) busca(s), que podem ser de formatos múltiplos (texto, áudio e vídeo). Cada máquina, naquele horizonte, que recebe a solicitação confere seu catálogo pessoal e os catálogos que existem, como metadados, dos demais por ela avistados em seus respectivos horizontes. Isso potencializa enormemente cada pesquisa. Recebidas as "respostas", o usuário escolhe de onde fazer o *download* desejado, acompanhando os *uploads* simultâneos, que podem estar sendo realizados de sua máquina por outros usuários.

Os documentos assim trocados são cópias fiéis do documento original, ficando agora disponibilizados também no computador daquele usuário remoto. A propriedade intelectual destes documentos (arquivos de música, vídeo ou texto) não está mais sob um controle centralizado. Mais ainda, estes protocolos agora permitem segmentar os documentos solicitados, fazendo o *download* simultaneamente de diferentes segmentos do arquivo desejado, de diferentes usuários, e imediatamente tornando estes segmentos disponíveis para outros *uploads* remotos (chamados de *leeches* – sanguessugas), o que aumenta geometricamente o número de fontes, para os demais, de onde fazer o *download*, bem como a dificuldade de impedir sua disseminação. Tudo isso com total privacidade e anonimidade dos usuários envolvidos, pois é virtualmente impossível descobrir quem está trocando o que e onde, destruindo assim muitas das velhas hierarquias de poder relacionadas com a informação e seu acesso.

Esta é uma das razões pelas quais tais protocolos foram destacados neste capítulo, pois o conhecimento de sua existência pode tornar totalmente fútil imaginar que os direitos autoriais, tal como existiram no século XX, possam ser preservados nesta nova mídia. Sempre e cada vez mais (à medida que usuários de computadores de uma geração nascida já sob a égide do digital crescem), haverá *hackers* dispostos a mostrar sua maestria e criar estratégias para atalhar os "obstáculos" legais que possam ser implementados.

Quando o *www* foi criado (a parte em hipermídia da Internet), as limitadas páginas existentes eram acessadas por meio dos *links* nelas colocados. Posteriormente, surgiram os indexadores, portais etc. Ia-se para um "lugar" para buscar a informação (muitos destes cada vez mais carregados de anúncios e que, basicamente, procuravam controlar o fluxo de informação na rede mediante da venda de posições privilegiadas no topo do catálogo personlizado para cada busca, utilizando uma indexação laica e anárquica).

A nova geração dos programas P2P ignora estes portais comercializados e os indexadores com suas limitações. Ela é desenvolvida por indivíduos ou pequenos grupos que se debruçam sobre as fragilidades ou deficiências das versões P2P anteriores, aprimorando-as mais e mais. Por exemplo, cada computador guarda agora um índice dos documentos avistados nas conexões anteriores, fazendo o processo de busca muito mais rápido e eficiente. Como dizem seus defensores:

> A noção de que propriedade intelectual é a mesma coisa que a propriedade de objetos físicos e tangíveis está entranhada na legislação atual [...] Infelizmente, historicamente, legalmente ou de qualquer outra forma, isto é simplesmente errado. (Eisgrau)[14]

O mais importante que as redes P2P fazem é tornar a colocar o fluxo da informação de volta nas mãos dos usuários, em oposição a uma economia do *copyright*[15] que tenta arrastar para esta nova mídia um modelo proprietário arcaico e inadequado, demonstrando com isso um compreensão falha do poder da informação em rede. O poder que isto representa é fantástico. Os dados que estão no HD dos usuários é que se mostram valiosos. Portanto, o que é compartilhado é aquilo que de valor se possui. Cada usuário controla este seu "compartilhamento", podendo decidir o que incluir ou excluir a cada momento. Este poder é realmente diferente do de compartilhar uma biblioteca digital numa *webpage* e merece uma séria avaliação por parte dos profissionais que estão conscientes da importância de tornar a informação disponível. E, tornando 'introdução deste capítulo, uma reflexão bem-informada e imparcial sobre o que é afinal direito "autoral", o que é direito "intelectual" e o que é direito "comercial", bem como da relevância e da "exequibilidade", se é que tal palavra

exista, de seu controle. Uma reflexão que parta do "reconhecimento de que cópias físicas de artefatos comuns, mas caros, possuem muito pouco valor, se uma cópia digital e acessível(ável) dos mesmos já existe"[16].

6. Outras questões

Ausência de *estandards* para descrição de páginas

O capítulo referente às *meta-tags* certamente irá discudir com mais profundidade este tópico. É importante mencionar "as dificuldades que essa ausência traz para a localização eficiente das informações e no impacto disso na atuação do profissional da informação, que tem agora de conhecer muito mais as ferramentas de pesquisa e o *how-to* da *web* do que propriamente as fontes de informação (ao contrário das fontes em papel). Isso é uma nova perspectiva para o bibliotecário, por dois motivos: primeiro, porque muda o perfil das suas competências, como comentei acima; segundo, porque pode ser um novo mercado de trabalho, como indexador, mesmo que seja exclusivamente para atender as demandas internas de sua instituição" [17].

Essa ausência de *estandards*, resultado da explosão de produção "laica" de documentos *on-line*, afastados do controle de editoras, bibliotecas e catálogos comerciais, de forma desorganizada e não-hierárquica, cria pouca diferença "perceptível" ao leigo entre documentos autênticos, fofocas, desinformação, notícia jornalística e pura especulação. Permite também, a quem distribui e/ou diponibiliza seus documentos no ciberespaço, criar estratégias "ocultas" para garantir melhor indexação e acesso, independentemente da autenticidade e da qualidade do conteúdo distribuído. O desafio de criar estratégias e metadados para garantir a confiabilidade e a adequada indexação de documentos em bibliotecas digitais e em indexadores comerciais, sua procedência de fontes fidedignas, diminuindo ou até evitando a anarquia atual dentro de um perspectiva ética e profissional[18], é mais uma razão para o profissional da ciência da informação conhecer os aspectos técnicos específicos da informação digital.

A volubilidade da Web

Atualmente a citação de documentos localizados na Internet é algo muito complicado devido à grande rapidez com que documentos são pu-

blicados, alterados e removidos. Coletâneas de *links* de um ou mais anos freqüentemente apresentam uma alta taxa de erro 404: avisos de documentos não mais existentes (ou que mudaram de "lugar").

Um das interessantes experiências atuais para administrar de forma organizada e legal alguns destes problemas é o *Open Archives* da Creative Commons[19], que pretende definir um espectro flexível de licenças que se acomodam entre o *copyright* pleno até domínio público, definidos pelo autor do documento disponibilizado. A iniciativa tem o objetivo de garantir a divulgação do conhecimento científico, que atalha sobre a burocracia da publicação em papel, agilizando a publicação web e, ao mesmo tempo, garantindo a solidez dessa referência, mediante dos compromissos dos repositórios em manter a informação disponível.

7. Encerrando o capítulo, em abril de 2005

Ao longo destes poucos anos de Internet aberta ao público não-acadêmico, diversos paradigmas foram propostos como substitutivos daquele(s) que representava(m) uma biblioteca "tradicional", quando da pré-existência deste volume explosivo de documentos digitais. Indiscutivelmente, as futuras aquisições da maioria destas bibliotecas irão incluir gradualmente mais e mais documentos digitais. Habilitar seu *staff* e seus usuários a utilizar, identificar, avaliar e recuperar estes documentos torna-se uma tarefa hercúlea, mas inadiável.

A tecnologia dos indexadores também evoluiu bastante ao longo destes poucos anos. Em suas versões mais atuais, como o Vivisimo[20] e o NorthernLight[21], eles analisam o universo encontrado, buscando as regularidades que permitam construir subcategorias que melhor auxiliem o usuário a encontrar o de que precisa, mostrando os mapas visuais dos critérios selecionados. Ajudar a aprimorar as versões subseqüentes também é um desafio para o profissional da informação.

A inexorável evolução da tecnologia quase que cotidianamente nos apresenta com novas alternativas, obsoletelizando as anteriores mais recentes. Projetos como o Liquid Information[22] mostram que o próprio hipertexto oportunizado pelo HTML (que superou com eficiência as limitações hierárquicas da tecnologia anterior, o *gopher*) vê-se, por sua vez, confrontado e talvez obsoletizado por projetos alternativos, que, num

futuro próximo, superem limitações como a de um *link* apontar para um único endereço ou documento. O conceito que defendem é o de que informação já foi analógica, já foi digital, e agora, pelo menos metaforicamente, quer ser "líquida", livre, fluída, maleável, personalizável. O formato digital a descolou do suporte. O formato "líquido", uma nova etapa em sua evolução, vai fazê-la existir em um fluxo ininterrupto e desobstruído. Isso porque, segundo afirmam seus criadores em seu *site*, a maioria das atividades relacionadas com a comunicação eletrônica preocupou-se com sua produção e com a disponibilização de seu acesso, não com as alternativas de processamento desta informação.

Então, o que permanece e o que muda no mundo digital? Esse é um assunto polêmico e controverso sobre o qual não há um consenso. Na desta autora, o mais importante norteador das decisões éticas, financeiras, educativas e políticas em relação a informação é torná-la disponível.

Se o usuário não encontra o que procura em um determinado lugar, volta-se para outras opções oferecidas pela rede e, se a biblioteca encontrar-se amarrada e amordaçada por conceitos tradicionais e obsoletos, outras fontes de informação equivalente surgirão rapidamente. Em poucos anos, o público jovem que precisará acessar as bibliotecas terá tido pouca experiencia com as originais e com seus artefatos curiosos e obsoletos, o livro impresso.

A luta entre a oligarquia e (o que é visto como) a anarquia digital na realidade não alteraram muito o ecossistema informacional digital, que cotidianamente se reorganiza no ciberespço, colaborativa ou individualmente criado, indiferente às limitações da mídia analógica. Estes ecossistema se desdobra, amplia-se e evolui de forma totalmente descentralizada, o que esperamos que tenha ficado claro ao longo deste capítulo.

Na Biblioteca do Congresso Americano, assim como no resto do mundo, comissões[23] ou indivíduos (professores, estudantes, profissionais da área) debruçam-se sobre as questões relativas às bibliotecas digitais e às mudanças em comportamentos e conhecimentos exigidas do profissional que as atenderão. Pode parecer frustrante e desalentador abrir mão de paradigmas arduamente aprendidos e testados e dispor-se a aprender novos fatos e habilidades, criando e testando hipóteses que refaçam ou substituam conceitos familiares, mas não mais adequados, e refazen-

do-os à medida que a tecnologia evolui. Indubitavelmente novos desafios se apresentam, alguns conceitos se obsoletizam ao mesmo tempo em que novos campos se abrem para o profissional da informação. Seu sucesso ou não parece ser, mais uma vez, diretamente proporcional ao domínio que tenham das características da informação: agora digital.

Notas

[1] Marcum, D The Future of Cataloging. [online] Address to the Ebsco Leadership Seminar Boston, Massachusetts (January 16, 2005) <http://www.loc.gov/library/reports/CatalogingSpeech.pdf> (tradução da autora)

[2] Guscott, J.(Editor of Library Futures Quarterl)These Emerging Technologies Will Change Public Libraries [online] (Updated May 1, 2001; originally posted February 1, 2001) http://www.libraryfutures.com/freereports/technology.htm

[3] LEVACOV, M. Bibliotecas virtuais [r]evolução?. Ciência da Informação, Brasília, DF, v.26, n.2, p.125-135, (maio/ago. 1997)

www.ibict.br/cionline/260297/26029702.htm

[4] Fineberg, G. Library Needs a Digital Strategy. [online] Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challanges of the Networked Resources and the Web (data?) http://www.loc.gov/catdir/bibcontrol/nas1.html

[5] Birdsall, W. The Myth of the Electronic Library: Librarianship and Social Change in America. Westport CT: Greenwood Press, 1994, p. xiii.(tradução da autora)

[6] http://www.ndc.uff.br/textos/vanja_periodicos.pdf

[7] Lombardi, J. Academic Libraries in a Digital Age. [online] D-Lib Magazine (October 2000) Volume 6 Number 10 ISSN 1082-9873 http://www.dlib.org/dlib/october00/lombardi/10lombardi.html

[8] ECO, Umberto. AFTERWORD. "Isso" (o livro) "substituirá aquilo" (a catedral), isso é, a informação filtrada e divulgada pela Igreja Católica. (tradução da autora) http://www.stanford.edu/dept/HPS/HistoryWired/Eco/EcoAfterword.html

[9] BARRETO, Aldo odla@centroin.com.br"o processamento técnico está acabando?" (07-03-2005)

[10] Marcum, D The Future of Cataloging. [online] Address to the Ebsco Leadership Seminar Boston, Massachusetts (January 16, 2005) http://www.loc.gov/library/reports/CatalogingSpeech.pdf

[11] DONIA, R. Death of the Book: An Historian's View of the Digital Revolution Rebuild [online] http://www.openbook.ba/icsl/papers/rdonia.htm

[12] SAFFO, P apud? Students shun search for information offline [online] (19-12-2004) (tradução da autora) http://www.filfla.com/newsitem.php?id=188

[13] http://p2p.libraries.psu.edu/

[14] "*The notion that intellectual property is the same as real or tangible property has become ingrained in current policy-making bodies and has become, if not dogma, at least accepted lore. Unfortunately, historically, legally, and otherwise, it's* just wrong". (tradução da autora) http://www.ala.org/al_onlineTemplate.cfm?Section=mwreports&Template=/ContentManagement/ContentDisplay.cfm&ContentID=59337#p2p

[15] Vaidhyanathan, S. P2P and the future of Information. The new information ecosystem: Part 1: cultures of anarchy and closure. [online] (27-06-2003)

http://www.opendemocracy.net/themes/article-8-1319.jsp#

[16] LOMBARDI, J. Academic Libraries in a Digital Age. [online] D-Lib Magazine (October 2000 Volume 6 Number 10 ISSN 1082-9873

http://www.dlib.org/dlib/october00/lombardi/10lombardi.html

[17] Grings, Luciana. email pessoal para a autora (20-04-05)

[18] http://www.capurro.de/diglib.htm

[19] http://creativecommons.org/

[20] http://Vivisimo.com

[21] http://www.northernlight.com

[22] What happens when you unleash text? http://www.liquidinformation.org/

[23] http://www.cetus.org/acad_lib.pdf

4. Gestão em bibliotecas digitais

Serviço de referência digital

*Marília Alvarenga Rocha Mendonça**

1. Introdução

Pode-se dizer que uma biblioteca tradicional é uma organização aberta, pois se acha inserida no meio ambiente que a cerca, influenciando-o e, ao mesmo tempo, sendo influenciada por ele, composta de funções e atividades relacionadas com a formação, desenvolvimento e organização de coleções (funções meio) e com a disseminação e recuperação da informação (funções-fim), produzindo produtos e serviços que satisfaçam às necessidades informacionais de seus usuários.

Decorrente deste meio ambiente que se transforma cada dia com uma rapidez cada vez maior e do propósito de atender às necessidades dos usuários, acompanham a mutação ocorrida na sociedade, imposta pelos avanços científicos e tecnológicos, a biblioteca tradicional convive com o surgimento de um outro tipo de biblioteca, a *biblioteca virtual*, conectada à rede e atendendo a uma gama de usuários superior à da biblioteca tradicional, visto não se prender a limitações impostas pelo tempo e espaço.

O surgimento da *biblioteca virtual* proporcionou uma transformação no *modus operandi* das bibliotecas, principalmente no tocante ao atendimento ao público, e o bibliotecário, intermediário entre o usuário

*Professora do Departamento de Ciência da Informação/UFF.
mariliaalvarenga@terra.com.br

e a informação, precisa dominar as técnicas bibliográficas juntamente com as novas tecnologias, para bem desempenhar seu papel neste novo cenário.

O serviço de referência, por ser o que proporciona esta interface usuário/informação, também sofre alterações, decorrentes do surgimento da *biblioteca virtual*. Surge o *serviço de referência digital*, que tem por finalidade prestar assistência aos usuários, virtualmente. Ferreira (2004, p.2) comenta que: "Fornecer um serviço de referência automatizado aos utilizadores da biblioteca e modernizar o serviço tradicional é um assunto complexo que os bibliotecários têm de enfrentar".

Refletir sobre o papel que os serviços de referência digital vêm desempenhando nas bibliotecas torna-se importante neste momento, visto a expansão que se delineia no cenário informacional. Conhecê-los e saber como funcionam é importante para avaliar se esse serviço cumpre o papel do serviço de referência tradicional. Fazer um registro da informatização das bibliotecas e relembrar as origens do serviço de referência e sua evolução é o que se pretende no momento.

2. As bibliotecas na era da informática

Lancaster (1994) faz referências a mudanças significativas, previstas desde 1960 nas atividades desenvolvidas pelas bibliotecas, advindas do emprego da tecnologia, principalmente nas de circulação, serviços técnicos e nas relativas à recuperação da informação através do assunto. Por outro lado, muitas das fontes de informação, base para o trabalho de referência, foram se transformando em base de dados informatizadas que, aliadas às novas ferramentas de busca disponibilizadas na rede, vieram transformar os serviços e produtos oferecidos pelas bibliotecas.

Segundo Briquet de Lemos (1998, p. 364), desde os idos de 1950, o computador começou a ser utilizado em serviços bibliotecários e de informação, como na "produção de listas bibliográficas, empréstimos de materiais e outras atividades de natureza gerencial" e, nos meados de 1960, os países desenvolvidos já produziam bases de dados bibliográficos que podiam ser consultadas a distância. O grande avanço se deu, no entanto, com o advento e, mais precisamente, com o rápido desenvolvimento da Internet a partir da década de 90, quando "possibilitou a

interconexão de computadores de diferentes marcas e com diferentes sistemas operacionais, utilizando linhas telefônicas comuns combinadas com linhas de transmissão de dados de alta velocidade".

As bibliotecas e demais unidades de informação se beneficiaram grandemente das facilidades oferecidas pela Internet, o que veio contribuir para facilitar a intermediação entre o fornecimento de informação e o usuário, possibilitando que as mesmas cumpram com maior eficiência seu papel na sociedade. As bibliotecas passam a atuar segundo novo paradigma, que privilegia a informação em detrimento do documento em si, o acesso ao invés da posse, não se limitando a local nem se prendendo a horários previamente determinados e que muitas vezes foram empecilhos para os usuários na obtenção da informação desejada.

Das bases de dados referenciais começam a surgir as textuais, oferecendo o texto completo dos documentos, juntamente com serviços de fornecimento de cópias cuja transação ocorre *on-line* (Comut) e que vêm ao encontro de uma das necessidades dos usuários de hoje em dia, ou seja, a otimização do tempo de espera pela informação desejada. Assim, os avanços tecnológicos vêm proporcionando mudanças consubstanciais nas bibliotecas: do modelo tradicional, com o acervo composto, basicamente, por material impresso e áudio-visual, para um acervo eletrônico e/ou virtual, resultando no surgimento das bibliotecas eletrônicas.

Cumpre ressaltar a opinião de Day e outros, citados por Rowley (2002), que afirmam existir uma variedade muito grande de significados a respeito de biblioteca eletrônica pelo fato de o referido termo não estar, ainda, consolidado na área. Confirmando esta opinião, Briquet de Lemos afirma que " na realidade ainda existe muita confusão a respeito do que se poderia chamar genericamente de bibliotecas eletrônicas". Assim é que são encontrados na literatura diversos termos para designar este tipo de biblioteca, destacando-se, entre eles bibliotecas sem paredes, bibliotecas em rede, bibliotecas no microcomputador, biblioteca lógica, biblioteca virtual, centro nervoso de informação (Corral, citado por Rowley, 2002).

Ao apresentar um significado para biblioteca eletrônica Oppenheim, citado por Rowley (2002, p. 4), "descreve-a como uma coleção organizada e administrada de informação numa variedade de meios (texto, imagem fixa, imagem em movimento, som, ou suas combinações), porém todos em formato digital", organizada de tal forma que permita o acesso

do para satisfazer as atuais necessidades informacionais de transição pelas quais as bibliotecas convencionais vêm passando". Assim é que:

> A biblioteca híbrida é designada para agregar diferentes tecnologias, diferentes fontes, refletindo o estado que hoje não é completamente digital nem completamente impresso, utilizando tecnologias disponíveis para unir, em uma só biblioteca, o melhor dos dois mundos (o impresso e o digital) (Garcez e Rados, 2002, p. 47).

Acredita-se que esta seja a situação da maioria das bibliotecas acadêmicas brasileiras, como está sendo delineado por uma pesquisa em andamento, realizada por Marcondes, Mendonça; Hughenin (2004), cujo objetivo geral consiste em averiguar se as bibliotecas universitárias brasileiras oferecem o serviço *referência digital* e em que consiste este serviço. Segundo Rusbridge, citado por Garcez e Rados (2002, p. 45), as bibliotecas híbridas devem proporcionar "uma vasta gama de interfaces, incluindo diferentes tipos e formatos de informação", integrados entre si, visando tornar todos os recursos acessíveis aos usuários, em que se destacam a existência de:

- OPAC (On-line Public Access Catalog) local (telnet/web);
- COPAC (Curl Public Access Catalog) – catálogo unificado *telnet/web*, isto é, participação em consórcios para utilização de recursos de outras bibliotecas;
- Catálogo regional virtual unificado;
- CD-ROMs e disquetes *off-line*;
- textos completos;
- grupos de dados remotos;
- grupos de dados locais;
- portais diversos (locais e remotos de recursos da web);
- jornais e livros eletrônicos;
- livros e jornais impressos;
- coleções especiais, mapas, slides, gravações de áudio e vídeos.

O acesso às informações nas *bibliotecas híbridas* se dá por meio, de uma adequada estratégia de busca, com a utilização de descritores e ca-

beçalhos que representem com presteza o tema pesquisado, exigindo "conhecimento do assunto, das bases de dados e da bibliografia" (Garcez e Rados, 2002, p. 50), com a ajuda do bibliotecário ou por um usuário devidamente preparado para tal.

O serviço de referência: das origens à referência digital

O serviço de referência é uma atividade que, segundo Maciel e Mendonça (2000), pertence à função encarregada pela *dinamização das coleções*,

> [...] engajadas no propósito final das bibliotecas, ou seja, o atendimento aos usuários reais e virtuais, a conquista dos potenciais e a decorrente dinamização do uso das coleções [...] Compreende todas as atividades voltadas, direta e indiretamente, à prestação de serviços ao usuário. Inclui a divulgação de informações gerais sobre a biblioteca [...] assim como as específicas, voltadas para um segmento específico. Promove o uso do sistema e de seus recursos e, para isso desenvolve atividades de treinamento pessoal de clientes na utilização do acervo, catálogos, bases e o acesso às bibliotecas virtuais.

Hutchins (1973) já sentia certa dificuldade em definir trabalho de referência, considerando como restritas algumas definições encontradas na literatura restritas (como as que consideram apenas o auxílio prestado aos usuários com os recursos da biblioteca), em contraponto a outras que considerava muito amplas, como no caso das definições que consideram que todas as funções desempenhadas pela biblioteca auxiliam os leitores.

A expressão *serviço de referência* apareceu, pela primeira vez, em um texto apresentado a uma conferência, por William B. Child, em maio de 1891, em que salientava a importância da experiência no trabalho do bibliotecário de referência (Grogan, 2001, p. 1) e que deixa clara a importância do elemento humano, em sua participação face a face com o usuário.

É bom lembrar que, até meados do século XIX, os estudiosos possuíam suas próprias bibliotecas e conheciam muito bem o acervo e o conteúdo das obras nelas contidas e, por este motivo, quase não consultavam outras bibliotecas. Por sua vez, "os bibliotecários, que amiúde também eram pessoas eruditas e conheciam intimamente o conteúdo dos acervos colocados sob sua custódia, seriam capazes de ajudar os leitores em suas pesquisas" (Grogan, 2001, p. 24). No entanto, este serviço não

mereceu uma posição de destaque nas bibliotecas que, naquele tempo, tinham as atividades de aquisição, catalogação, classificação e controle.

O acesso à educação por parte da população e o incremento da produção editorial, segundo Grogan (2001), contribuiu para mudanças nos serviços bibliotecários. Com a produção bibliográfica em expansão, com o crescimento de obras especializadas em diversos assuntos, o controle bibliográfico por parte dos eruditos ficou mais difícil, provocando uma demanda para a busca por assunto, "a que os bibliotecários responderam com mais catálogos de assuntos, sistemas de classificação e ajuda pessoal" (Grogan, 2002, p.24).

A democratização do ensino, por sua vez, gera novo público leitor que passou a exigir um novo tipo de biblioteca – a biblioteca pública mantida com impostos –, principalmente nas grandes cidades industriais da Grã-Bretanha e dos Estados Unidos e quando se pode localizar as origens daquilo que hoje conhecemos como serviço de referência. No entanto, continua Grogan (2001), apesar de em 1888 Melvil Dewey empregar a terminologia 'biblioteca de referência', o serviço de referência, como atividade rotineira das bibliotecas públicas, só veio a se consolidar no início do século XX.

Em se tratando de outros tipos de bibliotecas, as universitárias, por exemplo, o serviço de referência teve um impulso com o aliamento da pesquisa ao ensino, o que apontou para uma necessidade maior dos usuários pela utilização das fontes bibliográficas disponíveis nas bibliotecas, fato que, no entanto, só ocorreu nos meados de século XX. Já as bibliotecas especializadas foram consideradas "instituições do século XX, criadas sob medida para a finalidade de fornecer serviço de referência e informação" (Grogan, 2001 p. 28).

O serviço de referência é aquele realizado face a face com o usuário. Costuma-se dizer que é o *cartão de visitas* da biblioteca, responsável pela manutenção dos usuários, ou, como diriam os administradores, aquele que garante a fidedignidade dos clientes à empresa. Para que tal aconteça, é necessário prestar um serviço de qualidade, ou seja, um serviço que atenda às reais necessidades de seus usuários: a informação pertinente, no tempo certo, a um custo ótimo.

Para Hutchins (1973, p.4), "o contato direto com o leitor é o alvo do trabalho de referência". Ele acrescenta que:

> O trabalho de referência inclui a assistência direta e pessoal dentro da biblioteca a pessoas que buscam informação para qualquer finalidade e também as diversas atividades biblioteconômicas destinadas a tornar a informação tão acessível quanto possível.

Esta atividade que coloca o bibliotecário face a face com o usuário é muito semelhante ao de um professor, pois, além de fornecer informação, preocupa-se em "orientar pessoas na aquisição de conhecimentos e prepará-las para que possam, sozinhas, buscar informações sempre que precisarem" Mueller, citada por Martucci (2000, p. 104). Martucci vai mais além e consegue sintetizar o papel do bibliotecário de referência quando diz:

> O profissional de referência é aquele cuja atribuição no espaço informacional, especialmente nas bibliotecas, é o atendimento ao público. Seu fazer cotidiano é a interação com os usuários, chamada na área de processo de referência. Sua prática profissional é educativa, ao vivenciar com os usuários situações de troca, ao informar e ser informado, ao orientar o uso do espaço, dos instrumentos de recuperação e das fontes e recursos de informação, ao formar habilidades, valores e atitudes para o acesso, obtenção e uso da informação (Martucci, 2000, p. 103).

De acordo com Grogan (2001, p. 50), "a expressão processo de referência passou a ser empregada, nos últimos 30 anos, para denominar, em sua totalidade, a atividade que envolve o consulente e durante a qual se executa o serviço de referência", englobando desde a análise da natureza dos problemas do usuário até o fornecimento de informações capazes de solucionar estes problemas. Este processo envolve oito passos que se iniciam com o surgimento de um *problema*, susceptível de ser resolvido através da informação; a identificação da *necessidade de informação*, às vezes vaga e indefinida; a formulação de uma pergunta originária da questão apresentada pelo usuário; a *questão inicial* transformada na forma de uma pergunta; a *questão negociada* que consiste na reformulação da questão, caso haja necessidade de mais esclarecimentos para se chegar a uma solução adequada; a *estratégia de busca* que consiste em "uma análise minuciosa do tema em questão identificando seus conceitos e suas relações" para, em seguida, ser traduzido em "um enunciado de busca apro-

priado na linguagem de acesso do acervo de informações" (Grogan, 2001, p. 53); o *processo de busca* no acervo de informações; a *resposta*, que consiste no resultado da busca; a *solução* é a fase em que se chega à conclusão de que o resultado foi satisfatório. Todo este processo serve para lembrar que "[...] o processo de referência não se torna um espetáculo em que o bibliotecário é o único ator; deve ser um diálogo, com consulente e bibliotecário desempenhando papéis complementares do começo ao fim" (Grogan, 2001, p. 55).

3. O serviço de referência digital

O serviço de referência é uma das atividades da biblioteca que vem utilizando as facilidades oferecidas pela tecnologia, principalmente com o desenvolvimento da Internet e das inúmeras ferramentas que ela oferece, favorecendo, sobremaneira, a recuperação da informação pelas bibliotecas. Souto e outros (2004, p. 7) destacam que: "[...] a implantação de serviço virtual requer, além do domínio das ferramentas de comunicação, habilidades pessoais de comunicação e conhecimento dos recursos lingüísticos de interação eletrônica". Os serviços de referência digital podem ser considerados como uma evolução dos serviços bibliotecários via Internet e, cada vez mais, praticados pelas bibliotecas, já dizia Márdero Arellano (2001), pensamento compartilhado por Cuenca e outros (s.d.), ao se referir ao atendimento de questões de referências pelas bibliotecas acadêmicas e que, segundo ela, seriam as "pioneiras no uso das novas tecnologias para o acesso à informação".

De acordo com a opinião de Lankes, citado por Ferreira (2004, p.3): "Referência Digital diz respeito a uma rede de conhecimentos técnicos, intermediação e recursos colocados à disposição de alguém que procura informação num ambiente em linha".

Para Ferreira (2004, p. 3), a referência digital apresenta como antecedentes a própria biblioteca tradicional com seu serviço de referência e estudos realizados no campo da ciência da informação e os serviços "Ask A Services" e [...] "a utilização de sites de comunidades de peritos associados numa rede digital, que respondem a questões colocadas através da Internet".

O serviço de referência tradicional se caracteriza pela transação face a face com o usuário, quando as necessidades de informação dos usuários

podem ser identificadas. Pomerantz e outros (2002, p. 2) comentam que, desde a criação da Web, aumentou consideravelmente o número de serviços de referência que vêm utilizando meios de comunicação eletrônica assíncronos, para realizar este tipo de interação com os usuários. Os serviços de referência digital surgiram no final da década de 80, quando as bibliotecas começaram a disponibilizar seus catálogos na rede, segundo Márdero Arellano (2001), o que veio facilitar a localização de documentos de uma biblioteca.

No entanto verifica-se que, na prática, o serviço de referência digital oferecido pelas bibliotecas se refere, basicamente, em responder questões dos usuários que sentem necessidade de informação e que não sabem como encontrá-la por si mesmo, utilizando meios eletrônicos. Márdero Arellano (2001, p. 8) acrescenta: "Atualmente, muitos desses serviços estão reduzidos a consultas enviadas por correio eletrônico, telefone ou formulários na WEB, consumindo tempo e exigindo um trabalho árduo de pesquisa." Opinião semelhante apresentam Cuenca e outros (s.d.), ao se referirem a bibliotecas acadêmicas: "O atendimento de questões pela Internet tem sido uma nova opção para o usuário[...]".

Em recente pesquisa realizada nos Estados Unidos, cuja coleta de dados aconteceu no período entre fins de 2001 até início de 2002, por Pomerantz e outros (2004, p. 4), foi relatado que são poucos os modelos de referência digital que adotam procedimentos, como os empregados no balcão de referência tradicional. Informam ainda que, dentre os serviços de referência digital pesquisados, as práticas mais utilizadas eram o envio das respostas automatizadas para os usuários via e-mail e a manutenção de um formulário na Web para receber as questões dos usuários, inclusive, com espaço para informarem o endereço eletrônico, sem o qual não é possível retornar ao usuário.

Alguns serviços de referência digital armazenam as perguntas e respectivas respostas em um banco de dados de assuntos e, quando recebem uma pergunta, a base de dados é automaticamente acionada. Alguns desses serviços permitem o acompanhamento do processo registrando/informando a situação do mesmo, ou seja, se trata de uma nova questão, se recebida e ainda não respondida, se encontra em andamento. A pesquisa em base de dados das questões previamente respondidas, segundo Pomerantz e outros (2004), é tudo que serviços de referência digital desejam, apesar de ser importante a seleção das questões que podem ser res-

pondidas por meio desta ferramenta, ou, por outro lado, referem-se a questões que necessitam da mediação humana para serem respondidas. Informam ainda que alguns serviços mantêm os arquivos com as respostas previamente respondidas somente para os especialistas, ao passo que outros serviços os disponibilizam para todos os usuários.

Márdero Arellano (2001, p. 7), relata que existem, no exterior, inúmeros serviços de consulta que atendem a um volume grande de perguntas diariamente e

> [...] são operados por empresas atendendo à demanda de clientes e seguindo os interesses financeiros de seus investidores. Por outro lado estão os serviços de referência digital "Ask-A" não comerciais financiados por centros de pesquisa e usados por estudantes, pais, educadores e outros para se conectarem com profissionais experientes da área da biblioteconomia e serem guiados às melhores fontes de informação e especialistas de todas as áreas do conhecimento.

Como exemplo de alguns serviços de referência virtual, Márdero Arellano (2001) cita os seguintes:

- via correio eletrônico;
- Ask A Service, como o Ask-A-Scientist (recebem perguntas sobre uma área da ciência via e-mail);
- Ask-A-Librarian – acessado através da *homepage* da biblioteca.
- Ask-An-Expert – conta com a participação de especialistas (voluntários, orientadores, bibliotecários e outros) e" podem ser chamados de especialistas da informação porque realizam o trabalho similar àquele já estabelecido na biblioteca física" (Márdero Arellano, 2001, p. 8).

Márdero Arellano (2001, p. 7) relata que: "Ainda no início de 2001, nenhuma das 184 bibliotecas brasileiras cadastradas pelo IBICT no GT de Bibliotecas Virtuais (http://www.cg.org.br/gt/gtbv/alfabetica.htm) mantinha algum serviço de referência que funcionasse 24h e com distribuição seletiva do tipo de consultas". A pesquisa em andamento de Marcondes, Mendonça e Hughenin (2004) mostra, em análise preliminar de parte da pesquisa com bibliotecas universitárias brasileiras, alguns dos serviços/produtos que estão sendo oferecidos:

- Levantamento bibliográfico;
- COMUT;
- Pergunte à Bibliotecária;
- Fale conosco;
- Envio de cópias *on-line*;
- Sugestões;
- Empréstimo entre bibliotecas;
- Sumários correntes;
- Fichas catalográficas;
- Normalização de documentos;
- Dicionário, Enciclopédias *on-line*;
- Livros e periódicos *on-line*;
- Informativo: acervo, equipe e serviços;
- Lista de periódicos eletrônicos.

Uma questão que vem preocupando estudiosos desse tema está relacionada à mediação humana no processo de referência digital. O fator humano sempre foi considerado imprescindível no referido processo, utilizado pelas bibliotecas tradicionais e um dos responsáveis pelo sucesso das mesmas. Lankes, citado por Ferreira (s.d.), é de opinião que "é tempo de colocar um rosto humano na biblioteca virtual" e que o sucesso da biblioteca virtual depende das pessoas que nela atuam. Ferreira (s.d.) questiona: "Será que as TIC tornam obsoletas ou dispensável a mediação humana no Serviço de Referência?" Ferreira continua relatando: "Um outro problema formulado por Lankes relaciona-se com a falta de acompanhamento que os bibliotecários têm proporcionado aos utilizadores."

Torna-se importante definir o papel da biblioteca digital e do bibliotecário neste novo cenário. Pomerantz e outros (2004) se referem à diferença existente entre o atendimento a um usuário no balcão de referência de uma biblioteca tradicional, durante a entrevista de referência, em que o bibliotecário tem a oportunidade de tornar claras as necessidades de informação desse usuário, o que não é possível na referência digital, e a questão inicial apresentada pelo usuário é tudo que o bibliotecário tem para trabalhar. Informam ainda que serviços de referência digital mostram que, meios assíncronos não garantem a eles uma boa negociação. Como numa biblioteca tradicional, o planejamento desse serviço é indis-

pensável, nele incluído o estudo de usuário, sem o qual não é possível conhecer suas necessidades de informação e as formas de acesso por eles desejadas.

4. Considerações finais

O serviço de referência digital já é uma realidade nas bibliotecas e, no Brasil, convive com os serviços tradicionais de referência. Esse serviço encontra-se vinculado principalmente a bibliotecas acadêmicas, públicas e especializadas, como também a serviços independentes, sendo que diferentes serviços demandam práticas diferentes.

Observa-se que, no exterior, com destaque para os Estados Unidos, os serviços de referência digital são organizados com a finalidade de responder questões dos usuários, muitos deles gerenciados automaticamente e utilizando *softwares* aplicativos desenvolvidos para tal ou através de especialistas que são acionados para responderem a questões relativas a assuntos que dominam.

No Brasil, os serviços de referência digital podem ser encontrados através de *links* dispostos nas *homepage* das bibliotecas e/ou dos sistemas de bibliotecas que as abrigam (caso das bibliotecas acadêmicas). Nota-se uma semelhança com os serviços de referência de bibliotecas convencionais, apresentados em novo formato: o digital. Encontram-se, neste caso, as informações a respeito das bibliotecas (como acervo, serviços, equipe técnica, por exemplo), como também sobre os serviços oferecidos por outras instituições. Algumas bibliotecas permitem o acesso a documentos que se encontram em formato eletrônico/digital e, outras poucas oferecem o atendimento a questões de referência via correio eletrônico, como o caso dos serviços denominados "Fale conosco" ou "Fale com o bibliotecário".

É importante que o serviço de referência digital seja planejado, visando à sua estruturação ideal, compatível com o interesse dos usuários e a disponibilidade do sistema, com previsão de uma forma de avaliação periódica, visando a proporcionar um serviço de qualidade.

Quanto ao aspecto da participação humana no processo, acredita-se que continuará a existir, pois, mesmo em serviços oferecidos virtualmente, o bibliotecário se faz presente, seja quando disponibiliza informações

úteis a seus usuários, seja quando aponta as melhores fontes ou quando apresenta a resposta certa às questões apresentadas, fruto de seu conhecimento técnico especializado, aliado à sua habilidade de lidar com as ferramentas advindas das novas tecnologias e, principalmente, de sua experiência adquirida no desempenho profissional.

Acredita-se que a tendência é que a oferta dos serviços de referência digital se amplie nas bibliotecas e que se instalem serviços que funcionem em *tempo real*, utilizando-se dos *chats* e das *mensagens instantâneas*, entre outros recursos que a tecnologia possa oferecer.

Referências

BRIQUET DE LEMOS, Antônio Agenor. Bibliotecas. In: CAMPELLO, Bernadete Santos; CALDEIRA, Paulo da Terra; MACEDO, Vera Amarante, org. *Formas e expressões do conhecimento:* introdução às fontes de informação. Belo Horizonte: Escola de Biblioteconomia da UFMG, 1998. p 347-366.

CUENCA, Angela Maria Belloni e outros. *O usuário remoto de uma biblioteca acadêmica na área de saúde pública.* Disponível em <http://www.informacaoesociedade.ufpb.br/artigos/remotos.pdf> Acesso em: 01/02/05.

FERREIRA, Maria Izabel Goulão de Matos. *High Tech/High Touch*: serviço de referência e mediação humana. Disponível em:http://www.apbad.pt/pdf8Cong/com/com29.pdf. Acesso em 27/09/04.

GARCEZ, Maria Stuart; RADOS, Gregório J. Varvakis. Biblioteca híbrida: um novo enfoque no suporte à educação à distância. *Ci. Inf.*, Brasília, v. 31, n. 2, p. 44-51, maio/ago. 2002.

GROGAN, Denis. *A prática do serviço de referência.* Brasília: Briquet de Lemos/Livros, 2001.

HUTCHINS, Margaret. *Introdução ao trabalho de referência em bibliotecas.* Rio de Janeiro: FGV, 1973.

LANCASTER, F. W. Ameaça ou oportunidade? O futuro dos serviços de biblioteca à luz das inovações tecnológicas. *R. Esc. Biblioteconomia. UFMG*, Belo Horizonte, v. 23, n. 1, p. 7-27, jan./jun. 1994.

MACIEL, Alba Costa; MENDONÇA, Marília Alvarenga Rocha. *Bibliotecas como organizações.* Rio de Janeiro: Interciência; Niterói: Interciência, 2000.

MARCONDES, Carlos Henrique; MENDONÇA, Marília Alvarenga Rocha; HUGUENIN, Suzana. *Serviços de referencia digital em bibliotecas brasileiras.* 2004 (Projeto de Pesquisa submetido ao Programa Interinstitucional de Bolsas de Iniciação Científica – PIBIC/UFF-CNPq).

MÁRDERO ARELLANO, Miguel Angel. Serviços de referência virtual. *Ci. Inf.*, Brasília, v.30, n. 2, p. 7-15, maio/ago. 2001.

MARTUCCI, Elisabeth Márcia. Revistando o trabalho de referência: uma contribuição teórica para a abordagem interpretativa de pesquisa. *Perspect. Cienc. Inf.*, Belo Horizonte, v. 5, n. 1, p. 99-115, jan./jun. 2000.

MÁRDERO ARELLANO, Miguel Angel. Serviços de referência virtual. *Ci. Inf.*, Brasília, v.30, n. 2, p. 7-15, maio/ago. 2001.

POMERANTZ, Jeffrey e outros. The current state of digital reference: validation of a general digital reference model through a survey of digital reference services.

Information Processing & Management, v. 40, n. 2, p. 347-363, mar. 2004. Disponível em: http://www.sciencedirect.com/science?-ob=ArticleURL&-_udi=B6VC8-47PP6DR-1&-coverDate. Acesso em: 13/11/2004.

ROWLEY, Jennifer. *A biblioteca eletrônica*. Brasília: Briquet de Lemos/Livros, 2002.

SOUTO, L. F. e outros. Referência virtual: e-mail como ferramenta de interação com os usuários remotos de bibliotecas digitais. In: SEMINÁRIO NACIONAL DE BIBLOTECAS UNIVERSITÁRIAS, 13, 2004, Natal. Anais... Natal: UFRN, 2004. CD-ROM.

Gestão em bibliotecas digitais

Luiz Atilio Vicentini *

Estamos vivendo um período de transformações tecnológicas, culturais e sociais. Como estas transformações influenciarão as novas gerações, uma certeza, o acesso ao conhecimento será mais fácil e rápido, o aprendizado terá uma alavancagem substancial, comparada aos tempos atuais, aonde enfrentamos no dia a dia o assombro ao conhecimento registrado e divulgado com tanta rapidez, situação que não poderíamos imaginar até a 10 anos antes, principalmente para quem vem dos anos 50, 60 e 70.

Peter Druker (1995, p. 43) em seu livro "Administrando em tempos de grandes mudanças", resume bem o que significa estas transformações quando diz:

> Numa questão de décadas, toda a sociedade se rearranja – sua visão do mundo, seus valores básicos, suas estruturas sociais e políticas, suas artes, suas instituições básicas. Cinqüenta anos depois existe um novo mundo, e as pessoas nele nascidas não conseguem nem mesmo imaginar o mundo no qual viveram seus avós e nasceram seus próprios pais.

O cenário tecnológico aponta para a necessidade de mudanças que pode ser tão significativo quanto à mudança da arquitetura computacional do "mainframe" para o de cliente/servidor ocorrida na década de 80. A tecnologia atual vem crescendo através do desenvolvimento de sofistica-

* Coordenador do Sistema de Bibliotecas da UNICAMP. Especialista em sistemas de informação em ciência e tecnologia, pela PUCCAMP, e gestão de negócios e tecnologia de informação, pela FGV. vicentin@unicamp.br

dos sistemas de troca de mensagens, soluções de código aberto e novos protocolos de segurança, facilitando novos tipos de relações com maior colaboração e parcerias.

Situando as bibliotecas desse milênio, estas devem dirigir seus esforços em um novo conceito de estrutura, com a implantação de bibliotecas híbridas, trabalhando fundamentalmente na "logística da informação" armazenada, coletada e acessada. Segundo Davenport (1998),

> [...] a informação não pode ser considerada de maneira isolada nas instituições; e que, às bibliotecas está reservado o papel de repensar suas atividades e funções, adaptando-se aos novos modelos organizacionais e extraindo das tecnologias disponíveis o substrato para a melhoria na prestação de serviços e na utilização eficaz de informações.

A atual convergência digital conduz as bibliotecas, principalmente as universitárias, a buscarem mecanismos para atualização e recuperação eficiente de informações. O reposicionamento de mecanismos ágeis e eficazes tomou força a partir do advento da Internet, cujas transformações filosóficas e as novas formas de relacionamento advindas com essas novas tecnologias de informação lançaram luzes para maior difusão da informação científica registrada nas universidades.

Atualmente, a tendência mundial das unidades de informação é dispor seus acervos de forma eletrônica/digital, visando à conservação e/ou à disponibilização de seus conteúdos e o compartilhamento de recursos informacionais.

No âmbito de uma universidade que tem como princípio de atuação fomentar o ensino a pesquisa e extensão, a biblioteca digital torna-se uma ação proativa para que a comunidade acadêmica possa publicar seus trabalhos de forma rotineira, por meio de sistemas via WEB, difundindo o conhecimento, otimizando o fluxo de comunicação científica e reduzindo o ciclo de geração de novos conhecimentos.

A Internet destaca-se como uma grande vitrine para "consumidores de informação" cada vez mais sabedores de como se beneficiarem dessa tecnologia, no seu dia-a-dia, tanto no trabalho quanto no lazer. As principais tendências para atender às características exigidas pelo atual consumidor de informação são:

- auto-serviço: ele é auto-suficiente em seu caminho na busca pela informação;
- integração de tecnologias para facilitar o acesso à informação;
- navegação eficiente para localização da informação.

1. A biblioteca digital

O desenvolvimento de uma biblioteca digital parte de um movimento global em muitos setores da sociedade, com direcionamento ao trabalho virtual e serviços eletrônicos hoje constantes devido ao grande avanço das tecnologias de informação.

Inovação e difusão do conhecimento, esta é a essência de uma biblioteca digital. Com esses dois pensamentos, e com o auxilio de tecnologias que permitam a transferência do conhecimento, não importando a localização física do receptor, a estruturação de uma biblioteca digital propõem etapas a serem exploradas, etapas que influenciarão nas atuais funções das bibliotecas, desde a revisão das instalações físicas, passando pela aquisição e desenvolvimento de coleções, catalogação, classificação e indexação, principalmente no uso do computador como instrumento central de difusão da informação.

A criação de uma biblioteca digital deve atender às características como bem define Cunha (1999, p. 255):

> [...] o conceito biblioteca digital aparenta algo revolucionário, mas, na verdade, ele é resultado de um processo gradual e evolutivo. A introdução de processos digitais nos diversos serviços comumente existentes numa biblioteca está provocando impacto, com reflexos positivos e negativos, nas funções e serviços de uma biblioteca. Não existe uma estratégia única a ser empregada na implementação de uma biblioteca digital. As estratégias, tal como as bibliotecas nascem num determinado tempo e, obviamente, sofrem influências da cultura e das situações econômico-financeiras.

A informação como elemento principal da necessidade humana torna-se indispensável para o desenvolvimento da pessoa. No livro *A informação: análise de uma liberdade frustrada*, Xifra-Heras (1975, p. 27) afirma:

> a informação contribui diretamente para a propagação de conhecimento e, por conseguinte, para a formação dos indivíduos. Por isso se configura

como uma liberdade individual que se concretiza no direito do homem a emitir, expressar e receber informações.

A partir desta afirmação, um instrumento como uma biblioteca digital muito contribui para a difusão do conhecimento e para a formação social dos indivíduos.

Como principais componentes em uma biblioteca digital podemos citar:

Coleção/conteúdo
Recursos humanos Equipe multidisciplinar Capacitação
Padronização Metadados MARC Formato do arquivo digital Padrão de digitalização
Tecnologia *Hardware* Software Livre Proprietário Flexibilidade de desenvolvimento Facilidade de gerenciamento da coleção digital Linguagem de programação Utilização de protocolos de comunicação para importação e exportação de dados
Digitalização
Garantia de direito autoral
Preservação do documento digital

2. As tecnologias de informação

Uma das grandes alternativas atuais em tecnologias de informação para estruturação de bibliotecas digitais, é a utilização de *software* livre (código

aberto), viável para que as instituições desenvolvam soluções (serviços e produtos) com maior velocidade a um custo muito baixo.

O *software* livre pode ser definido a partir das características a seguir:

- O *software* livre pode ser utilizado, copiado e redistribuído livremente;
- O *software* livre é distribuído livremente junto com o código fonte;
- Alterações, melhorias, otimizações ou correções efetuadas são obrigatoriamente distribuídas gratuitamente na nova versão do *software*.

O *software* livre propõe total liberdade de criar e inovar. Na Internet os *softwares* estão disponíveis a "custo zero". Com o código fonte disponível e um pouco de conhecimento é fácil adequá-lo a cada necessidade.

Uma das características necessárias em uma biblioteca digital é a sua estruturação para integração e interoperabilidade das informações dos documentos digitais, por meio de metadados, possibilitando a captura e disponibilização dos dados e a sua integração com outros bancos digitais e/ou referenciais, utilizando instrumentos tais como:

- Protocolo Z39.50, ou versões mais novas;
- Protocolo OAI – *Open Archives*;
- extração e importação de dados em XML;
- extração e importação de dados ISO2709;
- formato de descrição MARC e suas variações.

3. Gestão da biblioteca digital

Deixando a parte os aspectos tecnológicos, o modelo de gestão de uma biblioteca digital, deve atender acima de tudo às características de cada instituição, podendo o modelo ser adaptado às reais necessidades de cada instituição, ou seja, totalmente centralizado, parcialmente descentralizado e totalmente descentralizado. Na escolha do melhor modelo, é necessário verificar se a tecnologia a ser utilizada atende a pelo menos dois dos modelos, ser uma tecnologia flexível que permita adaptações rápidas quando houver necessidade de alterações. Nunca pensar na criação de uma biblioteca digital com algum produto (*software*) com características proprietárias, aquele que só permite alterações se o próprio dono do *software* achar conveniente alterá-lo.

Criar uma biblioteca digital com foco no usuário. Saber quais serão os nossos usuários potenciais é fundamental no planejamento dessa biblioteca. Nunca se deve pensar que o seu principal usuário será o da sua instituição, pensar assim será um grande erro. A partir da estruturação de uma biblioteca digital com acesso via Internet, toda categoria de usuário terá acesso ao documento publicado, com variados interesses, pessoais, acadêmicos ou até por curiosidade. Nesse sentido, torna-se fundamental identificar qual conteúdo será trabalhado para construir o banco digital.

Estabelecer um planejamento para a criação da biblioteca digital que atenda às características a seguir: igualdade do documento digital com o documento impresso, acesso direto do usuário à informação publicada, revisão do fluxo de trabalho na biblioteca com os principais envolvidos, aplicar conceitos de qualidade visando a minimizar erros, flexibilidade e capacitação dos funcionários envolvidos em todo o processo da biblioteca digital, execução de estudos e planejamento financeiro para garantia de *software*, *hardware* e rede para acesso a biblioteca digital.

Uma estratégia para o bom desenvolvimento de uma biblioteca digital é a constituição de uma equipe multidisciplinar, com profissionais de diversas áreas da instituição, estabelecendo prioridades e metas, e o mais importante: todos devem estar envolvido no projeto da biblioteca digital a ser estruturada. Esta estratégia deverá abranger os aspectos de:

- recursos humanos – envolvimento de colaboradores de todas as áreas da instituição;
- recursos tecnológicos – delimitação das tecnologias a serem adotadas no desenvolvimento da biblioteca digital;
- recursos motivacionais – incentivo à equipe e colaboradores para o desenvolvimento do produto biblioteca digital.

Outro importante trabalho a ser realizado para estruturar uma biblioteca digital é a conscientização do autor do documento e das áreas geradoras desses documentos, mostrando a importância da divulgação desse conhecimento registrado. Este trabalho é contínuo, deve procurar mostrar ao autor que o acesso em linha ao documento digital, gratuito e completo amplia sua difusão, permitindo que o seu documento seja um verdadeiro instrumento para satisfazer a demanda requerida pelo usuário.

Também se deve atentar aos aspectos legais. O direito sobre o documento digital será sempre do autor, garantindo a ele o direito de veto da publicação do documento digital. Para facilitar esta questão, pode-se estabelecer prazos para a publicação, mediante formulário de autorização com as seguintes opções ao autor: opção 1 – autorizo a publicação do documento digital; opção 2 – autorizo a publicação do documento digital após dois anos da assinatura desta autorização; opção 3 – consulte-me após dois anos da data de assinatura desta autorização para publicação do documento digital. Com a indicação das opções 2 ou 3, o documento digital poderá ter restrição de publicação de dois a quatro anos.

Estabelecer procedimentos para a publicação do documento em meio eletrônico na biblioteca digital deve ser um objetivo a ser seguido. Os arquivos contendo o documento eletrônico deverão estar no formato com a extensão que identifique sua estruturação, tais como:

- Ponto doc (MSWord);
- Ponto ps (post-Script);
- Ponto pdf (Adobe).

Documentos que não estiverem no formato estabelecido e autorizado para publicação podem não ser aceito. Nesse caso, o correto é orientar o autor na conversão de seu arquivo/documento para um formato aceito, ou a própria biblioteca pode assumir esta responsabilidade. Nesse caso, a biblioteca deve realizar a conversão do arquivo segundo os padrões de qualidade estabelecidos, observando a igualdade deste se o documento já existir no formato impresso.

O conteúdo textual do documento deve (preferencialmente) ser salvo em um único arquivo, no formato autorizado. Também podem ser disponibilizados documentos que estivem em mais de um arquivo, os quais deverão ter seus arquivos nomeados na seqüência que as informações estão dispostas no formato original do documento, que pode ser o impresso. A numeração das páginas e o conteúdo devem estar em conformidade com o indicado no sumário do documento. Importante também eliminar, dos arquivos eletrônicos, possíveis chaves que restringem o acesso ao conteúdo a ser publicado.

Na formatação do documento eletrônico, levar em consideração que padrão único facilita no gerenciamento e na padronização, porém uma

biblioteca digital nunca deve restringir qualquer outro formato de documento a ser publicado. O formato mais usual atualmente é o PDF. Verifica-se que a tendência é a disponibilização de diversos tipos de documentos em diversos formatos, ou ainda a implantação do conceito de "cluster", um documento texto em formato PDF, com os anexos desse documento em outros formatos de vídeo e sons, anexados ao documento texto principal.

Um dos pontos relevantes ao se construir uma biblioteca digital é o acesso. A partir da sua implantação, se não houver restrições, o acesso será universal e trará grandes vantagens conforme as citadas a seguir:

- informação organizada;
- confiabilidade;
- métodos de busca mais abrangentes;
- sem limitações geográficas;
- sem limitações de disponibilização do número de volumes;
- preservação dos materiais impressos;
- informação com acesso hipertextual;
- apresentação de diversos formatos de documentos (texto, som, vídeo);
- integração, difusão e democratização do conhecimento.

4. Arquitetura da informação

Todo sistema de informação que possibilite a entrada e recuperação de dados exige aplicações voltadas a uma arquitetura para a organização de grandes volumes de informação, de forma que os usuários possam ter acesso rápido e preciso à informação desejada.

Um sistema bem estruturado deve atender às necessidades dos usuários, oferecendo um ambiente virtual organizado, facilitando a navegabilidade do visitante, com uma abordagem centrada no usuário, conduzindo a personalização das interfaces de comunicação no processo de recuperação e visualização dos conteúdos digitais.

O processo de construção de uma biblioteca digital se torna amplo e complexo, sendo necessário um conhecimento da estrutura de sua instituição e da forma como os documentos são gerados e tramitam pela instituição. Por outro lado, deve sempre ser levado em conta o desenvolvimento tecnológico, sendo necessário possuir, em seus quadros de

funcionários, técnicos de informática devidamente conhecedores de tecnologias para o desenvolvimento adequado das ferramentas de gerenciamento e acesso aos documentos publicados em formato digital.

Camargo (2004) define a estruturação de uma biblioteca digital como,

> Construir uma biblioteca digital envolve coleções de documentos digitais em vários formatos, mídia e conteúdo e, associados a isso, componentes de *hardware* e *software* que operam em conjunto com diferentes formatos de dados e algoritmos, várias pessoas, comunidades e instituições com objetivos, política e cultura próprias.

A organização da biblioteca digital, ou seja, a arquitetura da informação, conceitualmente, para Straioto (2002, p. 20):

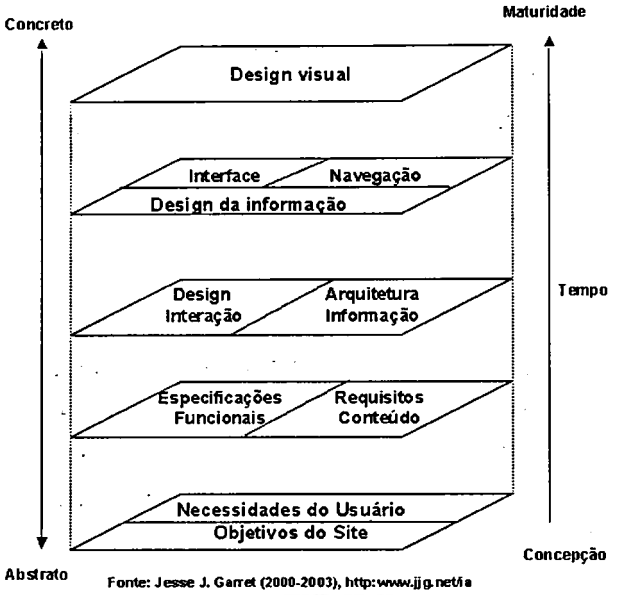
> [...] refere-se ao desenho das informações: como texto, imagens e sons são apresentados na tela do computador, a classificação dessas informações em agrupamentos de acordo com os objetivos do *site* e das necessidades do usuário, bem como a construção de estrutura de navegação e de busca de informações, isto é, os caminhos que o usuário poderá percorrer para chegar até a informação.

Em uma biblioteca digital este conceito também é válido. No momento da organização dos documentos digitais, deve-se levar em conta a estrutura da instituição detentora dos documentos. A organização pode ser feita por camadas ou tópicos, com nomes de áreas, departamentos, setores, facilitando inclusive o melhor entendimento da estrutura da instituição.

Uma arquitetura da informação bem elaborada compreende as etapas de organização, classificação e indexação do conteúdo, além da estruturação de canais de ligação para facilitar a navegabilidade entre as áreas da biblioteca digital e o acesso ao documento digital, construindo os caminhos da informação, suas conexões e desdobramentos, a fim de contribuir para a divulgação do conhecimento registrado e disponível.

A cultura digital atual nada mais é do que a adoção por parte dos indivíduos desta comunidade de novos padrões de interação e valores no mundo *on-line*. Torna-se necessário desenvolver projetos cada vez mais centrados na relevância e na praticidade para utilização, criando metodologias eficazes de organização e acesso à informação digital.

A fase de publicação de um documento em um banco digital transforma-se em um processo, com um conjunto de ações e envolvendo número significativo de pessoas.



Etapas de desenvolvimento de um produto WEB, com destaque para a fase da arquitetura da informação como elemento central do produto *on-line*

Deve-se ainda levar em conta os aspectos que envolvem o contexto tecnológico de um ambiente WEB e digital, sendo necessário à identificação detalhada dos elementos de um produto digital, a partir do ***hardware*** – quais e quantos deverão ser utilizados, ***software*** – quais sistemas operacionais aplicativos serão utilizados e desenvolvidos e finalmente o suporte digital – qual o tipo de suporte, armazenamento, local físico, mídia, condições ambientais previsão de reaproveitamento futuro.

O acesso aos documentos digitais deve ser livre, não deve existir restrição de visitas ou *downloads* aos documentos arquivados. O que pode ser adotado, o controle de *downloads* – O que o usuário da biblioteca digital está acessando? –, passa a ser uma exigência para execução de *downloads*. Esse controle pode ser feito por meio de um cadastro e autorização ao usuário que irá acessar e fazer *download* do documento digital. Com o emprego de um recurso no *software* gerenciador da biblioteca digital, o usuário realiza um breve cadastro, cria uma senha e, a partir desse momento toda vez que ele for acessar ou fazer *download* de algum documento com a se-

nha particular, fica registrado naquele documento quem acessou ou fez *download*. Importante, este banco deve ficar à parte do banco digital, e só pode ser acessado pelos administradores da biblioteca digital.

O acesso livre, navegabilidade, interface amigável, possibilidade de cópia de documentos sem custo, independentemente se existe ou não um controle de *downloads* aos documentos digitais, não inviabiliza o usuário na busca pelo conhecimento registrado nos documentos digitais. Todas estas facilidades atendem a um dos grandes requisitos da Internet, a usabilidade. Talvez uma das grandes restrições atuais é o tipo de conexão que o usuário a distância possui em sua estação de trabalho. Um usuário ao acessar a Internet por meio de uma conexão de baixa velocidade com certeza terá dificuldades em fazer um *download* de um documento digital, diferente do usuário que utiliza uma conexão de alta velocidade.

Um detalhe importante, no momento da criação de uma biblioteca digital, é o uso simultâneo de um mesmo documento. Dois usuários devem poder executar *downloads* de um mesmo documento simultaneamente sem prejudicar a *performance* do sistema.

Voltando à questão de quem acessa ou faz *downloads* nos documentos digitais. Essa é uma questão inusitada. De acordo com o conceito de uso na Internet, tudo é livre e o acesso é irrestrito. Na realização de um controle sobre *downloads* nos documentos digitais, o conceito de uso da Internet não é ferido, e o acesso continua livre e irrestrito, com uma informação a mais, e os administradores da biblioteca digital e principalmente a instituição poderão saber quem está e de onde está acessando os documentos digitais. Por meio desses dados, a instituição poderá ter indicadores de qual conhecimento registrado e divulgado pela biblioteca digital demonstra maior interesse. Essa informação poderá direcionar mais ou menos investimentos em determinadas áreas, por exemplo. Ao mesmo tempo, com esse tipo de indicador, passamos a ter uma nova biblioteca digital, uma biblioteca mais personalizada direcionada ao maior interesse da comunidade usuária.

Nesse ponto, retomamos à questão tecnológica. Muita das atividades descritas acima serão fáceis de serem implantadas, se o produto gerenciado (*software*) facilitar e tiver flexibilidade suficiente para adaptar-se à necessidade de gerenciamento não só do banco digital, mas também atender a mudanças necessárias de forma rápida e eficaz. Assim, é fundamental

que o *software*, ao ser desenhado, possua uma estrutura hierárquica que facilite a administração do fluxo dos documentos digitais, com atribuições bem definidas de "quem faz o que na biblioteca digital".

5. Diretrizes para o desenvolvimento e gestão de uma biblioteca digital

Para o desenvolvimento de uma biblioteca digital, deve existir a preocupação com aspectos relacionados à qualidade, consistência e credibilidade do conteúdo disponibilizado e certificação dos dados por meio da criação de vínculos automáticos para coleta de dados com bancos referenciais que já possuam as informações sobre os documentos registrados.

A estruturação de uma biblioteca digital pode seguir diretrizes baseadas em itens da teoria da qualidade, visando a estabelecer indicadores para que o projeto seja bem-sucedido, tais como:

Categorias	Variáveis	Características
Desempenho	Armazenamento dos Documentos	Integridade do documento digital; espaço em disco; controle de vírus; facilidade e rapidez na inserção; dados de conteúdo e informações adicionais; indexação dos arquivos; integração de dados.
	Mecanismo de Busca	Grau de recuperação dos documentos.
	Controle de Acesso	Registro de visitas e *downloads*.
	Cadastro de Visitantes	Identificação dos usuários que acessam as teses e dissertações.
	Estatísticas	Indicadores de visitas e *downloads*
Conformidade	Atores	Ações realizadas pelos personagens. Definição de quem faz o que?
	Arquitetura do Informação	Nível de organização das informações.
	Metadados	Padrões para a descrição dos dados.
	Integração e interoperabilidade	Utilização de protocolo de comunicação e protocolo de coleta automática dos metadados (*harvesting*); utilização de tecnologia de *software* livre.
Durabilidade	Implementações	Realização de *upgrade* de *software* e *hardware*.
	Segurança	Política de *backup*; documentação do *software* e da metodologia.
	Preservação	Política de preservação do documento digital.

A seguir são apresentados em detalhes às categorias e variáveis.

1) Desempenho – De acordo com Garvin (1992), "as características operacionais básicas de um produto é responsável pelo seu desempenho". No projeto de uma biblioteca digital, podemos identificar as variáveis mais importantes referentes a esta categoria:

a) Armazenamentos dos Documentos – Para o armazenamento dos documentos, o sistema deve contemplar em suas funcionalidades alguns requisitos de controle, necessários para garantia do desempenho da biblioteca digital:

- controle de *upload* – somente documentos autenticados são inseridos no banco digital;
- compressão dos arquivos – objetiva a economia do espaço ocupado pelos documentos, garantindo a *performance* de acesso ao banco digital;
- identificação de vírus – o servidor deve possuir um sistema de antivírus ativado pelo controle de *upload*, garantindo a integridade do arquivo inserido, do banco digital e dos *downloads* a serem realizados posteriormente;
- publicação do documento – a inserção do documento digital deve ser de fácil e rápida operacionalização. Buscar a integração da biblioteca digital com os bancos referenciais (catálogo automatizado), que, por meio de protocolos de comunicação (ex.: Z39.50), permite a transferência dos metadados, evitando o retrabalho na descrição dos documentos no banco digital;
- metadados – a descrição dos documentos na biblioteca digital é realizada mediante a marcação (identificação) dos metadados que são extraídos de outras bases referenciais.

b) Mecanismo de Busca – O mecanismo de busca é uma das partes mais importantes do sistema. O sistema deve manter uma base de dados própria otimizada para realizar as buscas. Essa característica garante a *performance* no tempo de resposta quando da submissão de uma pesquisa, independentemente da quantidade de documentos no banco digital. O sistema recupera palavras encontradas tanto nos campos indexados, quanto no conteúdo do documento.

c) Cadastro de Visitantes – O acesso aos documentos é liberado após o preenchimento do formulário de registro, ocasião em que o visi-

tante se identifica com seus dados, criação de login e senha, automaticamente autenticada pelo sistema.

d) Controle de Acesso – Possibilidade de registro de todas as visitas e *downloads* de cada arquivo, com a identificação do usuário que está executando o *download*. Após o primeiro *download*, o sistema permite ao usuário a realização de novos *downloads* sem a necessidade de nova autenticação da senha. Esse controle de acesso permite o reconhecimento dos documentos mais visitados.

e) Estatística – Permissão do registro do número de visitas e *downloads* realizados por documento. Apresentação dos dados estatísticos dos documentos mais acessados, por área do conhecimento, por domínios, instituição e pelos termos (assuntos) mais utilizados para as buscas no banco digital.

2) Conformidade – Essa categoria da qualidade é definida em Garvin (1992) como "o grau em que o projeto e as características operacionais de um produto estão de acordo com padrões preestabelecidos". A premissa básica da organização de qualquer acervo documental é a classificação dos assuntos por área do conhecimento. Outro exemplo ainda de instrumento normalizador são os cabeçalhos de assuntos, os vocabulários controlados ou ainda os *thesaurus*, utilizados para a indexação em bases e sistemas de recuperação da informação como requisitos de conformidade necessários ao estabelecimento de especificações técnicas para a padronização dos dados, que garantam a qualidade do fluxo e da recuperação das informações pelos seus usuários:

a) Usuários – O gerenciamento das ações desenvolvidas na biblioteca digital, mediante a da definição dos possíveis papéis que os usuários (internos e externos) possam ter dentro do sistema. A atribuição dos papéis dos personagens (administrador, responsável, colaborador e visitante) permite a definição das atividades que serão realizadas no ambiente da biblioteca digital de maneira sistêmica e padronizada, permitindo a descentralização das tarefas de maneira a garantir a autonomia controlada de cada papel desempenhado. O sistema deve prever quem faz o que, quando e como?

b) Arquitetura da Informação – O sistema deve possuir uma estrutura hierárquica para a organização das informações que serão definidas em níveis (1° nível, 2° nível etc.). Para cada especificidade dos assuntos, podem-se criar outros níveis, ou seja, um nível pode conter um ou mais

sub-níveis, dependendo da necessidade. Podem ser previstos também a definição de categorias que correspondem a tipos de documentos válidos e os limites de tamanho, podendo ainda especificar um ou mais formatos a serem aceitos dentro de cada categoria.

c) Metadados – Os metadados constituem-se em uma nova forma de estruturar e disponibilizar informação eletrônica por meio da Internet. A literatura identifica formatos para descrição de dados bibliográficos de acordo com as recomendações estabelecidas pela norma ISO 2709 e também dos metadados. Inclui informações descritas tradicionalmente, como os catálogos de bibliotecas (autor, título, assunto etc.) e sobre recursos eletrônicos. Gils (2001) relaciona a idéia de metadados às citações bibliográficas, entretanto seu conceito torna-se mais amplo em razão das suas características: descrever, identificar e definir o recurso eletrônico; apresentar diferentes níveis de especificidade, estrutura e complexidade; objetivando modelar e filtrar o acesso, estabelecer termos e condições para o uso, autenticação e avaliação, preservação e interoperabilidade das publicações eletrônicas.

d) Integração e Interoperabilidade – Esses itens são identificados nas ações de implementação de sistemas de acesso, disseminação e cooperação de repositórios informacionais, convergem para a possibilidade de os usuários interagirem em recursos informacionais heterogêneos, armazenados em diferentes servidores na rede, em uma interface única para a recuperação das informações demandadas.

3) Durabilidade – Diz respeito à vida útil do produto e, de acordo com Garvin (1992), "pode-se definir durabilidade como o uso proporcionado por um produto até ele se deteriorar fisicamente". As variáveis e características identificadas nessa categoria da qualidade da biblioteca digital são analisadas a seguir:

a) Novas implementações – O projeto da biblioteca digital deve prever a médio e longo prazo a realização de *upgrade* no *software* e *hardware*, objetivando manter a qualidade do sistema, independentemente do crescimento do seu banco digital.

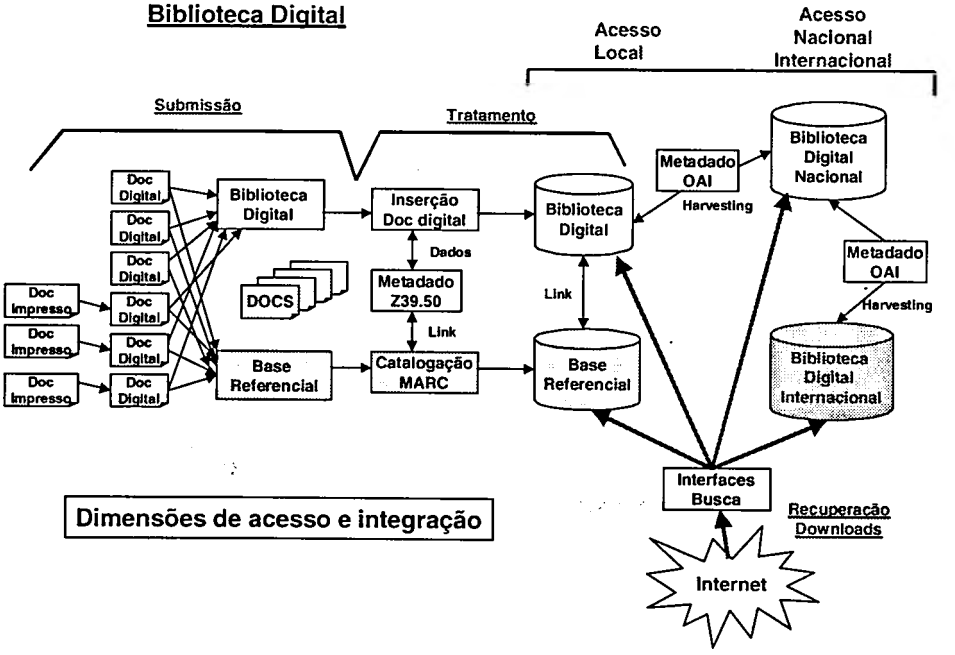
b) Segurança – A adoção de políticas de *backup* constante visando a garantir a integridade física dos documentos digitais, e a documentação do *software* e da metodologia da biblioteca digital são requisitos funda-

mentais para garantir a continuidade no seu desenvolvimento e as manutenções que se fizerem necessárias. Todo projeto baseado em tecnologia de *software* livre permite o desenvolvimento de novas funcionalidades e o aperfeiçoamento do sistema a medida de suas demandas.

c) Preservação do documento digital – Não existe ainda uma política de preservação do documento digital, em formato e mídia adequada para garantir o acesso contínuo dos materiais digitais para gerações futuras. Essa preocupação tem sido objeto de discussões pelos pesquisadores da área da informação e pelos responsáveis dos repositórios que utilizam a Web como ambiente informacional.

6. Uma biblioteca digital integrada

Para total interoperabilidade de uma biblioteca digital, esta deve estar integrada nos recursos de conteúdo, *hardware*, *software* e metodologia de publicação. Nesse sentido, apresenta-se a seguir um esquema gráfico demonstrando a dimensão da integração necessária para uma biblioteca digital:



A visão de futuro, apontada por diversos especialistas, acena para um acervo eletrônico/digital, em que o conceito do acesso deve predominar sobre o de posse da informação.

Uma nova biblioteca coloca-se no atual estágio de desenvolvimento tecnológico, a biblioteca digital não deve mais se preocupar somente com o desenvolvimento de sua coleção local visando a satisfazer às necessidades dos usuários presenciais. Os bibliotecários devem estar atentos a uma nova realidade na biblioteca, que terá no futuro um enorme acervo digital. Com essa mudança, será necessário conhecer os reais interesses dos usuários, principalmente o usuário não-presencial.

Devemos pensar na biblioteca digital do futuro, com uma estrutura mais personalizada e funcional. Personalizada no sentido de conhecer o seu usuário, o que foi mais acessado, de onde é o usuário, e criar mecanismos de interatividade com este usuário, por exemplo: criando um canal de comunicação informando os novos documentos publicados na biblioteca digital sobre o assunto por ele pesquisado anteriormente. Funcional no sentido de que a sua infra-estrutura contemple os requisitos de *hardware*, *software* e arquitetura da informação, para permitir maior flexibilidade para mudanças e adaptações exigidas em nível institucional e de usabilidade.

Referências

BIBLIOTECA Digital: publicação de teses e dissertações. Sistema de Bibliotecas da Unicamp. Procedimento interno SBU/P003/04.

CAMARGO, L.S.A. de; VIDOTTI, S.A.B.G.; CAMARGO, V.V. de. *Arquitetura da informação para Bibliotecas Digitais*: uma abordagem centrada no usuário. Disponível em: http://libdigi.unicamp.br/document/?code=8285. Acesso em: 20 jul. 2004.

CUNHA, M. Desafios na construção de uma biblioteca digital. Brasília, Ciência da Informação, v. 28, n. 3, p. 255-266, set./dez. 1999.

DAVENPORT, T. H. *Ecologia da Informação*: por que só a tecnologia não basta para o sucesso na era da informação. Tradução Bernadette Siqueira Abrão. São Paulo: Futura, 1998. 316 p.

DE ROSA, C.; DEMPSEY, L.; WILSON, A. *Análisis del entorno de 2003 por OCLC*: reconocimiento de patrones. Dublin (Ohio): OCLC, 2004. 174 p.

DIGITAL libraries and virtual workplaces: important initiatives of Latin América in the information age. Johann van Reenen, editor. Washington, DC, IACD, 2002. 226p.

DRUCKER, P.F. *Administrando em tempos de grandes mudanças*. São Paulo: Pioneira, 1995. 230 p.

GARVIN, D.A. *Gerenciando a qualidade*: a visão estratégica e competitiva. Tradução Engº João Ferreira de Souza. Rio de Janeiro: Qualitymark, 1992.

GILS: overview-ideas behind the GILS approach. Disponível em: <http://www.gils.net/overview.html>. Acesso em: 25 jan. 2005.

GONÇALVES, M.A.; FOX, E. Technology and research in a global networked University Digital Library (NUDL). Brasília, *Ciência da Informação*, v. 30, n. 3, p. 13-23, set/dez. 2001.

HEEMANN, V. Mudança de hábito: impacto das novas tecnologias na qualificação do profissional bibliotecário e no uso final. In: SEMINÁRIO SOBRE AUTOMAÇÃO EM BIBLIOTECAS E CENTROS DE DOCUMENTAÇÃO, 5., 1994, São José dos Campos, *Anais*... São José dos Campos: UNIVAP, 1994, p.172-176.

LODDO, M.N. *Biblioteconomia e arquitetura da informação*. Disponível em: <http://webinsider.uol.com.br/vernoticia.php/2185>. Acesso em: 14 jan 2005.

MARCONDES, C.H.; SAYÃO, L.F. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. Brasília, *Ciência da Informação*, v. 30, n. 3, p. 24-32, set/dez. 2001.

PAVANI, A.M. A model of multilingual digital library. Brasília, *Ciência da Informação*, v. 30, n. 3, p. 73-81, set/dez. 2001.

PORTAL de Tesis Electrónicas. Universidade do Chile. Disponível em: <http://www.cybertesis.net/acerca.html>. Acesso em: 06 jan. 2005.

QUEIROZ, R. *Software livre e inovação*. Disponível em: <http://www.comciencia.br/200406/reportagens/11.shtml>. Acesso em: 13 jul. 2004.

REENEN, J. Working in a digital world; new leadership practices for the development of a digital library workforce. Brasília, *Ciência da Informação*., v. 30, n. 3, p. 82-91, set/dez. 2001.

SPONCHIADO, R.A.; VICENTE, V.S. Biblioteca digital de teses e dissertações do Instituto de Física Gleb Wataghin – UNICAMP: relato de experiência. In: SIMPÓSIO INTERNACIONAL DE BIBLIOTECAS DIGITAIS, 2. 2004. Campinas: UNICAMP, 2004. Disponível em: http://libdigi.unicamp.br/document/?code=8263. Acesso em: 28 fev. 2005.

RAMOS, M.E.M. (Org.). *Tecnologia e novas formas de gestão em bibliotecas universitárias*. Ponta Grossa: UEPG, 1999. 257p.

STRAIOTO, F. A. *Arquitetura da informação para a world wide web:* um estudo exploratório. 2002. 120 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília.

TRISKA, R.; CAFÉ, L. Arquivos abertos: subprojeto da Biblioteca Digital Brasileira. Brasília, *Ciência da Informação*, v. 30, n. 3, p. 92-96, set/dez. 2001.

VICENTE, G.; BETTA, C.E.D.; SOUZA, K.S.; SIMÕES, D.F. O uso do protocolo de comunicação Z39.50 e a extensão biblioteca YAZ da linguagem de programação PHP e seus vínculos entre a Biblioteca Digital da Unicamp, tópicos dissertações e teses, TCCs, partituras e mapas digitais que utiliza *software* livre e o *software* integrado de funções VIRTUA. In: SIMPÓSIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 13., 2004, Natal. *Anais*... Natal: UFRN, 2004. 1 CD-ROM.

VICENTINI, R.A.B. *Análise dos padrões e indicadores de qualidade para disponibilização das teses e dissertações na Biblioteca Digital da Unicamp.* 2004. 99 f. Dissertação (Mestrado em Gestão da Qualidade) – Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, Campinas.

VIDOTTI, S.A.B.G.; SANCHES, S.A.S. *Arquitetura da informação em WEBSites*. Disponível em:

<http://libdigi.unicamp.br/document/?code=8302>. Acesso em: 25 jan. 2005.

XIFRA-HERAS, J. *A informação:* análise de uma liberdade frustrada. Rio de Janeiro: Lux, EDUSP, 1975. 346p.

ZAFALON, Z.R. Biblioteca Digital X Biblioteca Virtual: aspectos norteadores para proposta de implantação em uma IES. In: SIMPÓSIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 13., 2004, Natal. *Anais*... Natal: UFRN, 2004. 1 CD-ROM.

Bibliotecas digitais: uma nova cultura, um novo conceito, um novo profissional

*Helena Pereira da Silva**
*Othon Jambeiro***
*Ângela Maria Barreto****

1. Introdução

Discorrer sobre as competências do profissional da informação no planejamento e operação de bibliotecas digitais implica inseri-las, necessariamente, nas discussões sobre a profunda mudança de uma cultura analógica para uma cultura digital. Essa mudança está atrelada à eterna relação técnica e cultura, verificada ao longo do desenvolvimento da humanidade.

Sobre essa relação, Capurro, em palestra no 6° Congreso Nacional de Bibliotecologia y Documentación da Colômbia, em 2000 – cujo tema foi "Cultura Digital e Informação no Próximo Milênio" – lembra Ortega y Gasset, no livro *Meditación de la Técnica* (1965, *apud* Capurro, 2002, p.1), que distingue a técnica em três estágios: a natural; a do artesão; e a do técnico. Esta última sempre tem sido a meta do homem, nela inserida

*Doutora em engenharia de produção (UFSC), mestre em ciência da informação (IBICT/UFRJ), professor adjunto do Instituto de Ciência da Informação, UFBA. helenaps@ufba.br
**PhD in communication studies (University of Westminster, U.K), mestre em ciências sociais (USP), professor titular do Instituto de Ciência da Informação, UFBA, pesquisador. C do CNPq. othon@ufba.br
***Doutora em comunicação (USP), mestre em ciência da informação (PUC/SP). Professor adjunto do Instituto de Ciência da Informação, UFBA. ambar@ufba.br

a possibilidade de agir à distância, como coloca Capurro: uma *action in distans*, demonstrada pelas invenções que unem o homem com o distante: como a pólvora, a bússola, a imprensa etc.

A relação técnica e cultura é sustentada pelos processos comunicacionais. Ao produzir teias de significação para si mesmo, o homem fundamenta sua vida social nos aspectos da produção, armazenamento e circulação da informação e do conteúdo simbólico. Para isto, utiliza-se de vários recursos ou meios técnicos. Thompson (1998), ao inserir esta compreensão em sua discussão sobre a mídia e a modernidade, diz que os meios técnicos empregados necessitam apresentar algumas características, tais como as de fixação, reprodução e distanciamento temporal.

O grau de fixação depende do meio utilizado e liga-se aos mecanismos de armazenamento da informação. O atributo da reprodução refere-se à possibilidade de multiplicação de formas e conteúdos simbólicos. A terceira característica, o distanciamento espaço/temporal, significa o afastamento da forma simbólica de seu contexto de produção no tempo e no espaço. Alterações nestas características redundam na passagem de uma cultura analógica para uma cultura digital.

Sabe-se que todos os meios técnicos têm relação com os fatores espaço e tempo da vida social. A imprensa, por exemplo, é uma técnica que colocou o indivíduo – que tinha sua capacidade comunicacional fortemente sincrônica e espacialmente bastante restrita – em contato com essa periferia infinita – em espaço e tempo – de leitores potenciais, como sugerido por Ortega (idem, ibidem). O surgimento de novas técnicas, como o telégrafo, o telefone, o disco, o rádio, a TV, a era das telecomunicações, enfim, tornou o atributo tempo/espaço ainda mais significativo.

Capurro transporta essa idéia para a perspectiva da Internet e a mudança do paradigma cultural provocada pela técnica das redes digitais, que não só conectam o indivíduo solitário com uma periferia infinita de leitores, como permitem a comunicação desses leitores com o autor. Isto transforma ambos em participantes virtuais de um diálogo e de uma ação comum, que ele denomina *actio digitalis in distans* e – o que revoluciona tudo – de forma interativa em tempo real.

Outra possibilidade que a Internet oferece é a da mestiçagem cultural, um dos mais fascinantes fenômenos do mundo atual, tanto nos aspectos positivos de diferenciação e mútuo enriquecimento como nos ne-

gativos de homogeneização, incluindo, às vezes, políticas do tipo protecionista ou isolacionista. Com essas proposições, Capurro invoca o sonho Ortegiano, que se faz realidade agora:

> Desde hace años sueño com um posible curso em que se muestren frente a frente las técnicas de Occidente y las técnicas del Ásia (Ortega, 1965, p.95, *apud* Capurro, 2002, p.2)

Capurro (2002, p.2) complementa que a cultura digital é algo assim como um projeto vital em que a técnica do técnico, cristalizada na figura do engenheiro, muda em direção ao que se chama engenharia do conhecimento, podendo operar conjuntamente com a técnica do espírito, que pode ser representada em termos menos idealistas, como técnica de *software*. Não no sentido estrito, mas *latu senso*, para indicar uma forma de estruturar, desenvolver e comunicar o saber que oferece possibilidades específicas de uma *actio digitalis in distans*, não realizável dessa maneira em nenhum outro meio.

Estamos, sem dúvida, num contexto em mutação, onde já não são mais apenas previsões as mudanças significativas dos papéis profissionais tradicionalmente estabelecidos, em particular daquele que era o profissional da biblioteca: o bibliotecário. Elas já são reais. No seu papel tradicional, a função era estar fisicamente num espaço físico específico, processando, armazenando e recuperando documentos, suportes materiais, que o leitor pessoalmente buscava. Esperava pelo leitor numa atitude reativa e dentro de um contexto previsível. Utilizava-se de códigos convencionais na tarefa de mediar a informação ao usuário. Agora, é preciso considerar que a utilização dos novos meios técnicos para fixação e transmissão da informação exige novas formas de habilidades, competências e formas de conhecimento, pois empregam novas regras de codificação.

Os programas de Qualidade Total, desenvolvidos para empresas, terminaram por invadir também o mundo das bibliotecas, despertando os bibliotecários para uma atitude mais pró-ativa. Ainda assim, era a cultura do analógico, em que se previa o atendimento de necessidades de usuários "ao alcance das mãos e dos olhos". A chegada dos meios digitais e, com eles, da possibilidade de "esticar os braços" para muito além

do espaço físico em que atuavam, mudou o foco dos profissionais do documento para a informação.

Agora, quando o relacionamento com o usuário se dá via sistemas automatizados, um novo cenário da informação se configura, incorporando diferentes atores. É momento de redefinições, reinvenções e reposicionamentos do profissional e das chamadas bibliotecas digitais. Urs (2002, p.1) afirma que não se trata de uma mudança determinada apenas pelas tecnologias, mas também pelas tendências do mercado da informação, profundamente alterado por mudanças ocorridas na cadeia produtiva da informação.

Transformações em muitas áreas da sociedade se somaram à ascensão das TICs, foram intensas e atingiram todos os segmentos da sociedade do pós-guerra. Nas empresas, os formatos organizacionais tradicionais, fundamentados em forte centralização de poderes, funcionalização, hierarquização e comunicação verticalizada, além de excessiva burocratização das tarefas, remodelaram-se.

Na atual sociedade, as organizações de trabalho apresentam-se com grandes e fortes deslocamentos no que se refere à mudança de paradigma, aos aspectos econômicos – que têm dimensões mundiais – e à descentralização de poderes. Neste contexto, as informações tomam lugar privilegiado na arena econômica e começa-se a falar em unidades de informação:

> o negócio das unidades da informação é o de disseminar informações precisas e prestar correta assistência a empresas, para que iniciem, mantenham e concluam as necessárias mudanças tecnológicas, comerciais e gerenciais, logísticas e mercadológicas, que assegurarão sua evolução e sobrevivência (Ramos, 1996, p.16).

Vê-se impor à expansão histórica das bibliotecas um novo momento, aquele em que essa antiga instituição se posiciona articulada com os serviços das unidades de informação, o novo setor das organizações encarregado das funções de gerenciamento deste insumo básico de suas atividades. Amplia-se, pois, o papel do profissional da informação, que passa a desempenhar tarefa administrativa, além das técnicas, visto que a informação é entendida sob a ótica mercadológica.

Neste ponto é importante salientar a forma repetitiva como aparece a palavra "mudança". Ela acontece em cadeia e é o que permanece nesse contexto de rápidas transformações tecnológicas. Isto é, se o permanente é a mudança, a capacidade de adaptação deve também permanecer. Deve, na verdade, se transformar num programa de existência, como alerta Capurro (2003), o que constitui, hoje, um dos maiores desafios da humanidade.

Assim como mudança, desafio é uma palavra recorrente em textos que tratam da relação entre bibliotecários, bibliotecas e TICs. Urs (2002, p.1) aponta que as veneráveis instituições bibliotecas e seus profissionais estão sendo desafiados a oferecer novos serviços e novos atores diante desse novo mercado da informação. Portanto, qualquer programa de educação para bibliotecas digitais deve considerar não só as mudanças tecnológicas, mas, principalmente, as tendências que essas mudanças provocam no mercado da informação. O autor relaciona algumas que necessariamente devem ser consideradas:

- aos olhos dos clientes (usuários) as bibliotecas no ciberespaço são apenas mais um fornecedor no mercado da informação;
- há muita oferta de recursos e serviços de informação nesse mercado;
- há uma grande fragmentação desse mercado;
- as mudanças tecnológicas são mais rápidas do que se pode imaginar;
- as demandas e expectativas dos clientes (usuários) crescem a cada dia;
- cada vez mais, ampliam-se as necessidades de se oferecer serviços de informação de valor agregado.

Dessa forma, o mundo do trabalho se revolucionou com as TICs. O próprio significado de trabalho tem registrado importantes mudanças. O restrito conceito de *posto de trabalho* deu lugar ao conceito mais amplo de *ocupação*, entendida como atividade que não se restringe a um grupo de tarefas aglomeradas e operacionais ou a funções. Ela deve ser entendida como um conjunto mais aberto, que evoca os conhecimentos básicos de uma área, com a característica de poderem ser transferidos do exercício de uma para o de diversas outras ocupações.(Ferreira H. e outros, 2004)

A transformação do *posto de trabalho* em *ocupação* representa, na verdade, mudanças distintas: da consideração de trabalhador para indivíduo; do trabalho solitário para a valorização da socialização, com o traba-

lho em equipe; da orientação do esforço físico para o do cérebro (Vargas Zúñiga, 2000; 2001; Arruda, Marteleto, Souza, 2000). O perfil profissional exigido nestes novos tempos de intensa utilização de tecnologias de informação e comunicação prende-se ao tripé **conhecimentos, habilidades e atitudes**. Tal perfil não se restringe aos profissionais da informação, de maneira geral, nem especificamente aos tradicionais, como os bibliotecários. Abrange, na verdade, todos os que exercem algum tipo de atividade laboral, porque expressa as profundas mudanças do mundo do trabalho, o que implica também a necessidade de mudança na relação trabalho-educação, como colocam Arruda, Marteleto e Souza (2000, p. 15).

Volta-se, então, à mudança da cultura analógica para a digital, em que também se insere essa relação trabalho-educação. Na verdade, estamos diante de um novo *ethos* ético e sociopolítico, irrecusavelmente inserido no "espírito do nosso tempo", no qual também se insere o que se denomina sociedade da informação, baseada nas redes digitais, ou sociedade do conhecimento, que privilegia o saber perante o fazer. Ambas as denominações são unificadas pelo condicionamento do sucesso pessoal e social à aprendizagem contínua, numa conformação de círculo virtuoso e, por conseqüência, de uma educação global.

O conceito de *ethos*, já utilizado por Homero e Aristóteles, de acordo com Lastoria (2001, p.63), significa a morada do homem, isto é, a Natureza. Uma vez processada mediante a ação humana, sob a forma de cultura, ela faz com que a regularidade própria aos fenômenos naturais seja transposta para a dimensão dos costumes de uma determinada sociedade. A cultura promove, então, a sua própria ordenação, ao estabelecer normas e regras de conduta, que devem ser observadas por cada um de seus membros.

Segundo Caniello (2003, p.31-32), o conceito de *ethos*, apropriado pelas ciências sociais por Kroeber e Geertz, tem estreita identidade com a noção de *Volksgeist* (espírito de um povo), elemento central da filosofia de Hegel, para quem, o "espírito de um povo", ou cultura, estará sempre traspassado pelo "espírito do tempo" (*Zeitgeist*), que lhe impõe limitações e constrangimentos. Na síntese de Caniello (2003, p.32) sobre o pensamento de Kroeber e Geertz, o *ethos* guarda a marca da estrutura que conforma a tradição de um povo, seu "espírito", mas também comporta os influxos da ação dos sujeitos e das pressões conjunturais que interagem

com essa estrutura, em um determinado tempo histórico. Estamos, portanto, irremediavelmente, inseridos na cultura digital e adaptar-se a ela é seguir o espírito do tempo.

2. O conceito

Pensar sobre conceitos é um exercício intelectual essencialmente acadêmico e sempre dependerá da cultura vigente. O conceito tradicional de biblioteca, criado num "tempo de simplicidade", como coloca Peek (1998, *apud* Watstein; Calarco; Ghaphery, 1999), provavelmente não foi difícil de ser estabelecido. Etimologicamente, biblioteca é uma coleção de livros. Conceitualmente incorpora vários sentidos, mas basicamente é uma coleção de livros organizados a partir de uma estrutura que classifica o conhecimento e inserida em um determinado ambiente físico.

Além disso, carrega uma imagem construída ao longo do tempo: o sacrário do saber, que numa ambiência austera convida ao recolhimento do pensar. Um ambiente muitas vezes assustador, com seus códigos indecifráveis aos não iniciados, identificando peças sagradas e dispostas em "vielas", muitas vezes mal iluminadas e propícias a encontros com "fantasmas" zelosos do silêncio do seu espaço. Cunha (1999, p.257) relembra também a imagem do poder pela retenção do conhecimento, explorada por Umberto Eco em *O Nome da Rosa*.

A biblioteca tradicional é uma instituição secular e, sem dúvida, de grande importância social. Seu conceito é prático: coleções organizadas e encerradas entre paredes. É como ponte que liga o universo de usuários e o vasto mundo das idéias. Tem como função a coleta, organização e disseminação de informação, geral e/ou especializada, no âmbito de uma comunidade específica. A ênfase está sempre no edifício ou na coleção e não no sentido de biblioteca como uma organização, como afirmam Watstein; Calarco; Ghaphery (1999, p.346).

A mesma facilidade de construção não se aplica ao conceito de biblioteca digital. Desde a popularização do "ser digital" de Negroponte (1995, *apud* Watstein; Calarco; Ghaphery, 1999), as coleções de informações disponíveis no meio digital vêm recebendo uma atenção cada vez maior. No entanto, a nomenclatura e o conceito ainda suscitam controvérsias, como pode ser constatado no recente trabalho de Silva; Sá; Furtado (2004).

A diversidade de denominações e conceitos é discutida por Watstein; Calarco; Ghaphery, (1999) no artigo Digital Library: Keywords. Os autores relembram que, já em 1978, Lancaster (p.348) previu a sociedade sem papel (*paperless society*), observando a intensa migração dos processos informacionais para os computadores, desde alguns pioneiros, como o MEDLARS (Medical Literature Analysis and Retrieval System), em 1965, e a importância da rapidez da recuperação nesses sistemas, que em 1975 já contavam mais de 500 serviços.

Na verdade, essa previsão de mudança dos repositórios de informação remonta à Vannevar Bush, considerado o precursor da biblioteca digital. Ele previu esse futuro, em 1945, e apresentou o Memex: "um dispositivo em que o indivíduo armazenará seus livros, seus registros, suas anotações, suas comunicações. O dispositivo será mecanizado de modo a poder ser consultado com extrema velocidade e flexibilidade" (Bush, 1945, *apud* Silva; Sá; Furtado, 2004. 1)

As previsões de Bush se confirmaram com possibilidades talvez não imaginadas por ele. No entanto, o papel não desapareceu como preconizava Lancaster. Ao contrário, aumentou o volume em função da facilidade de edição e impressão. Mas, cada vez mais, a informação sobre as atividades humanas e o estoque informacional migram para o meio digital. Muitas informações já não são mais produzidas ou disseminadas via papel, como é o caso da infinidade de títulos de periódicos *on-line* (sem versão em papel) que estão surgindo. E deve-se salientar: muitas delas com acesso gratuito. Isso representa uma significativa mudança cultural, dado o rompimento do monopólio das editoras milionárias, em direção a tão propalada democratização do acesso ao conhecimento.

Em recente entrevista, Miguel Ángel Márdero Arellano, técnico do Ibict responsável pelo curso de Editoração Eletrônica de Revistas Científicas, explicou por que cresce, a cada dia, o número de organizações interessadas em disponibilizar suas revistas na Internet. As que já existem no formato papel estão migrando para a rede e novas estão sendo criadas diretamente no formato digital:

> Há uma tendência que aponta para a migração das publicações para o formato eletrônico. Evidentemente, a publicação em papel jamais desaparecerá porque ela tem um propósito para as bibliotecas, para o intercâm-

bio de informações. Entretanto, é bom que se note que a divulgação eletrônica está tendo uma divulgação muito maior. Especificamente no Ibict, nós temos verificado que a revista *Ciência da Informação* começou com uma certa resistência por parte dos usuários, mas agora o trabalho está sendo bastante divulgado. Além disso, nós podemos obter estatísticas e contabilizar não somente os acessos à revista *Ciência da Informação* como um todo, mas também constatar que artigo foi mais ou menos acessado. Outro aspecto muito importante é que aumentou o número de autores que estão submetendo seus trabalhos e também cresceu o número de acessos ao Portal do Ibict, com a divulgação da revista *Ciência da Informação* no sistema. (2005, disponível em http://www.ibict.br/noticia.php?id=82) (grifo nosso)

A fala de Arellano confirma a expressão corrente de que a facilidade do acesso amplia o potencial de acesso e uso. Além disso, os repositórios automatizados ou bibliotecas digitais permitem os estudos bibliométricos, informétricos e cienciométricos de forma ágil e muito mais segura do que os tradicionais métodos, que demandam muito tempo e imprecisões, por serem manuais. Esses estudos, que permitem "descobrir o conhecimento" por meio do cruzamento e análise de dados, ganham importância cada vez maior. Estão estreitamente relacionados aos indicadores da atividade científica e tecnológica, que se encontram no centro dos debates, na perspectiva das relações entre o avanço da ciência e tecnologia, por um lado, e o progresso econômico e social, por outro, como aponta Macias-Chapula (1998, p. 134).

Esse autor chama a atenção para o fato de que as revisões de políticas científicas pareceriam inconcebíveis, hoje, sem recorrer aos indicadores existentes. Se, por muito tempo, o foco das avaliações permaneceu orientado para medir os insumos, como verbas e pessoal de P&D (pesquisa e desenvolvimento), crescentemente o interesse está se voltando para os indicadores de resultados (Okubo, 1997; King, 1987, *apud* Macias-Chapula, 1998, p. 134).

Dessa forma, a informetria, de acordo com Tague-Sutcliffe (1992, *apud* Macias-Chapula, 1998, p.134), vem ganhando destaque mais recentemente entre os cientistas da informação na Europa e Estados Unidos, como um campo geral de estudo que inclui as áreas mais antigas da bibliometria e da cienciometria. A bibliometria, termo propagado por

Pritchard, em 1969, é o estudo dos aspectos quantitativos da produção, disseminação e uso da informação registrada. Desenvolve padrões e modelos matemáticos para medir esses processos, usando seus resultados para elaborar previsões e apoiar tomadas de decisão.

A cienciometria é o estudo dos aspectos quantitativos da ciência enquanto uma disciplina ou atividade econômica. É um segmento da sociologia da ciência, sendo aplicada no desenvolvimento de políticas científicas. Envolve estudos quantitativos das atividades científicas, incluindo a publicação e, portanto, sobrepondo-se à bibliometria. A informetria é, portanto, o estudo dos aspectos quantitativos da informação em qualquer formato, e não apenas registros catalográficos ou bibliografias, referente a qualquer grupo social, e não apenas aos cientistas. Pode incorporar, utilizar e ampliar os muitos estudos de avaliação da informação que estão fora dos limites tanto da bibliometria como da cienciometria. Uma tipologia para as três disciplinas foi identificada por McGrath (1989, *apud* Macias-Chapula, 1998, p. 135), com seus objetos de estudo, variáveis, métodos e objetivos dispostos no quadro 1.

Quadro 1 – Tipologia para definição e classificação da bibliometria, cienciometria e informetria

Tipologia	Bibliometria	Cienciometria	Informetria
Objetos de estudo	Livros, documentos, revistas, artigos, autores, usuários	Disciplinas, assunto, áreas, campos	Palavras, documentos, bases de dados
Variáveis	Número de empréstimos (circulação) e de citações, freqüência de extensão de frases, etc.	Fatores que diferenciam as subdisciplinas. Revistas, autores, documentos. Como os cientistas se comunicam.	Difere da cienciometria no propósito das variáveis; por exemplo, medir a recuperação, a relevância, a revocação, etc.
Métodos	Ranking, freqüência, distribuição	Análise de conjunto e de correspondência.	Modelo vetor-espaço; modelos booleanos de recuperação, modelos probabilísticos; linguagem de processamento, abordagens baseadas no conhecimento, tesauros
Objetivos	Alocar recursos: tempo, dinheiro, etc.	Identificar domínios de interesse. Onde os assuntos estão concentrados. Compreender como e quanto os cientistas se comunicam.	Melhorara a eficiência da recuperação

Fonte: Macias-Chapula (1998, p. 135)

O conceito de biblioteca digital deve envolver, então, também esse sentido de "celeiro de conhecimentos", a partir do cruzamento de dados numa rede de conceitos: dados, informação, inteligência, conhecimento – tomada de decisão – direcionamento de políticas sociais, econômicas e de informação. Pacheco; Kern, 2001, *apud* Silva; Sá; Furtado (2004, p.6) apontam a criação da Biblioteca Digital Brasileira – que teve como pioneiro o repositório de teses e dissertações do Programa de Engenharia de Produção da Universidade Federal de Santa Catarina – como um sistema único, que permite o acesso à produção científica, além de fornecer dados bibliométricos, informétricos e a medição de acesso.

Watstein; Calarco; Ghaphery (1999), tratando do conceito e da polissemia, fizeram um levantamento em três bases de dados nas áreas das ciências da informação e da computação, no período 1970-1997. Usaram as bases Library Literature, Inspec e Compendex, verificando o período e o número de artigos que aparecem sob cada palavra-chave. Chamam a atenção para o fato de que esse estudo é comparativo entre os diversos termos, diferentemente da maioria dos estudos encontrados na literatura até aquela época, que tratam dos termos de forma independente, sem compará-los.

A figura 1 demonstra a quantidade de artigos no período 1973-1986, no qual as curvas demonstram claramente a situação dos termos *paperless library* (biblioteca sem papel) e *library of the future* (biblioteca do futuro), os primeiros a surgir. São seguidos, a partir de 1980, por *on-line library* (biblioteca "em linha") e a partir de 1984 por *library without walls* (biblioteca sem paredes).

A figura 2 demonstra o período de 1987 a 1997, quando o termo *on-line library* se mantém, mas em declínio até 1997. Em 1987 surgem os termos *datawarehouse (depósito de dados)*, *virtual library* (biblioteca virtual) e *eletroctronic library* (biblioteca eletrônica). Em 1990 começa a ascender, de forma discreta, o termo *digital library* (biblioteca digital), que cresce numa curva quase vertical de 1994 até 1996, com um leve declínio em 1997, mas é o termo que permanece em maior evidência até o momento.

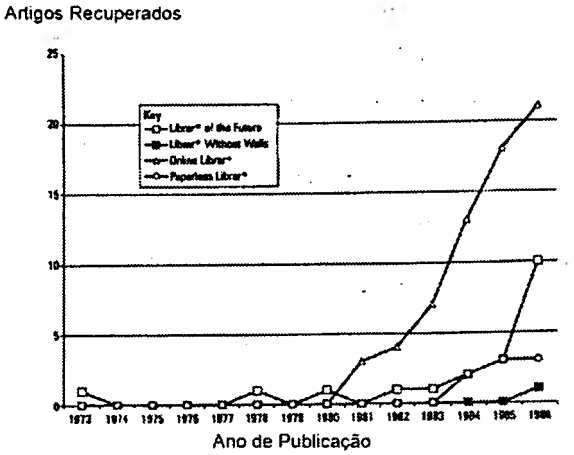


Figura 1 – As raízes das palavras-chave
Fonte: Adaptado de Watstein; Calarco; Ghaphery (1999, p.345)

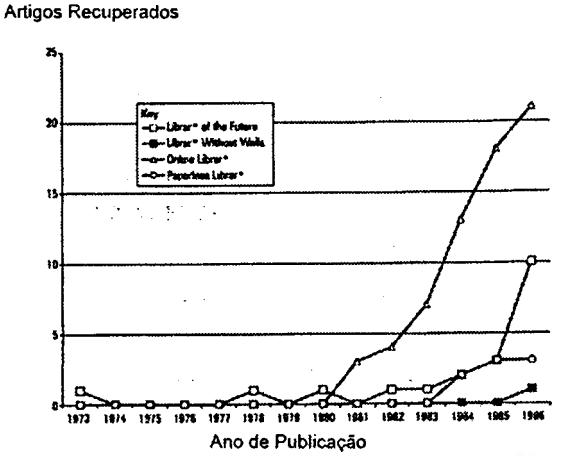


Figura 2 – A predominância dos termos *digital library* e *datawarehouse*
Fonte: Adaptado de Watstein; Calarco; Ghaphery (1999, p.346)

Na literatura nacional, trabalhos recentes demonstram que as indefinições quanto a terminologia e conceito permanecem. Martins (2002) no artigo *Perspectivas para uma biblioteca no futuro: utopia ou realidade* aborda os seguintes termos: Biblioteca tradicional; biblioteca eletrônica; biblioteca eletrônica virtual; biblioteca polimídia; biblioteca interativa; biblioteca virtual; biblioteca de realidade virtual; biblioteca digital e biblioteca universal.

Silva; Sá; Furtado (2004) ainda abordam a diversidade de terminologias que se referem aos repositórios informacionais no meio digital e a crescente importância do tema, a partir do desenvolvimento acelerado da Internet, desde 1980. Citam que há uma vasta literatura, tanto nacional quanto estrangeira, com inúmeras denominações para conceitos distintos, e que as expressões mais recorrentes são biblioteca virtual e biblioteca digital. Registram os trabalhos de Cunha (1997) e Ohira (2001) como exemplos de artigos nacionais que narram levantamentos bibliográficos sobre o tema. E apontam para o fato de que várias comunicações são feitas nos eventos da área da ciência da informação e biblioteconomia, além do *Workshop* Internacional Políticas de Informação em Bibliotecas Digitais. O próprio trabalho foi apresentado ao Simpósio Internacional de Bibliotecas Digitais em 2004, em Campinas-SP.

As mesmas autoras (Silva; Sá; Furtado, 2004, p. 2), optaram pelo termo biblioteca digital e explicam que o tema envolve todos os aspectos de construção de um banco de dados, com assuntos novos e de abrangência interdisciplinar. O estudo delas se concentrou em levantamento nos periódicos *Ciência da Informação e Datagramazero – Revista de Ciência da Informação* e nos *anais do Seminário Nacional de Bibliotecas Universitárias* (SNBU); do *Congresso Brasileiro de Biblioteconomia e Documentação* (CBBD); e do *Integrar – Congresso Internacional de Arquivos*. Os artigos encontrados foram agrupados em quatro eixos temáticos: 1-Conceitos; 2-Tratamento da Informação e Metadados; 3- Biblioteca Digital Brasileira (BD) e a Biblioteca Digital de Teses e Dissertações; 4- Implantação de Bibliotecas Digitais.

É significativo o fato de o resultado de um trabalho tão recente (2004), baseado na análise do uso do conceito por autores da área de ciência da informação, apontar para a presença ainda de imprecisão. As autoras citam como provável razão para o não-consenso a multidisciplinaridade, que envolve desde a concepção até a implantação da biblioteca digital. Importante salientar que o estudo destaca o fato de que a formulação do conceito está quase sempre nas ferramentas e nos processos, sem menção aos atores envolvidos, como os produtores de conhecimento, os gerenciadores – profissionais de informação e analistas – e os usuários. (Silva; Sá; Furtado, 2004,p.3). (grifo nosso)

A exemplo de Watstein; Calarco; Ghaphery (1999), fizemos uma consulta na base Scielo[1], exatamente no índice de assuntos, e encontramos

os seguintes termos dispostos em ordem alfabética: biblioteca digital; bibliotecas digitais; biblioteca digital multilíngüe; biblioteca digital universitária; biblioteca do futuro; biblioteca eletrônica; biblioteca híbrida; biblioteca virtual. Numa busca por cada um desses termos foram recuperados 41 artigos, no período 1997-2004. Todos são da revista *Ciência da Informação*, do IBICT, única revista da área indexada na base, até o momento da pesquisa (abril de 2005). A distribuição dos artigos sob cada termo ao longo do período está demonstrada no Quadro 2.

Quadro 2 – Termos e número de artigos recuperados no índice de assuntos da Base Scielo 1997-2004

TERMOS	1997	1998	1999	2000	2001	2002	2003	2004	Total
Biblioteca do Futuro	1			1					2
Bibliotecas Brasileiras na Internet	1								1
Biblioteca Eletrônica	4	1		1					6
Biblioteca Virtual/Eletrônica	1			1					2
Bibliotecas Brasileiras na Internet	1								1
Biblioteca Híbrida						1			1
Biblioteca Virtual	7			1		1			9
Biblioteca Digital ou Bibliotecas Digitais	2		1	2	9	2		2	19
Total Geral									41

A demonstração do quadro 2 é a recuperação pelo termo do índice de assunto. No entanto, feita a avaliação dos títulos, verifica-se a inconsistência do índice de assunto, ou da imprecisão do conceito e da terminologia. Buscando ainda constatar a utilização do termo **biblioteca digital**, o mais utilizado atualmente, fizemos o levantamento desse termo diretamente na base da revista *Ciência da Informação*, no mesmo período do levantamento da base Scielo: meados de abril de 2005. Utilizou-se o mecanismo de busca avançada e apenas no campo assunto. O resultado, apresentado no Quadro 3, a exemplo da base Scielo, aparece entre 1997 e 2004, no entanto com resultado de 14 artigos, contra 18 da base Scielo.

Quadro 3 – Termos e número de artigos recuperados no índice de assuntos da base da revista Ciência da Informação do Ibict - 1997-2004

TERMOS	1997	1998	1999	2000	2001	2002	2003	2004	Total
Biblioteca Digital ou Bibliotecas Digitais	2		1	1	7	2		1	14

Pelo exposto, está claro que a inconsistência permanece entre autores e linguagens de indexação. Nesse ponto é preciso retomar a questão da mudança cultural que necessariamente envolve os atores, foco do nosso interesse aqui. As outras categorias apontadas pelas autoras, que envolvem o tratamento da informação no contexto digital, as ferramentas tecnológicas e o estabelecimento de padrões, fundamental para a interoperabilidade necessária à implantação das bibliotecas digitais e o compartilhamento de dados, são essencialmente técnicas e necessitam de equipes multidisciplinares. E esse aspecto envolve novamente a mudança cultural. O trabalhar em equipe de forma sinérgica é um desafio para profissionais que sempre procuraram estabelecer domínios sem interação de forma efetiva.

Dada esta realidade, e numa tentativa de síntese conceitual e de posicionamento neste trabalho, assume-se que bibliotecas digitais são repositórios de informação disponíveis nas redes (referenciais ou de texto completo, de imagens fixas ou móveis e de sons), acessíveis de forma remota, de qualquer lugar do planeta. Essa condição está estreitamente relacionada com o conceito de recuperação da informação, desenvolvimento de ferramentas para esse fim, assim como ao estabelecimento e adoção de padrões que permitam a intercambialidade. O acesso por meio da vitrine global em que se transformou a Internet está tornando real o sonho de Paul Otlet e Henri LaFontaine, acalentado desde 1893, quando organizaram uma conferência para propor o Controle Bibliográfico Universal. (Taylor, 2004, p.29).

Essas perspectivas apontam para a possibilidade de uma biblioteca digital universal, cuja construção, por via da interligação de bibliotecas digitais temáticas, nacionais, locais, etárias, étnicas, escolares etc. depende de algumas condições (*core areas*), cujo preenchimento requer competências específicas dos profissionais de informação. Isto será discutido a seguir.

3. Áreas essenciais (*core areas*) da biblioteca digital: bases para as competências profissionais

As considerações sobre bibliotecas digitais e profissionais da informação, de acordo com as proposições encontradas, parecem estar vincu-

ladas à identificação dos aspectos essenciais que devem ser considerados nesse novo tipo de biblioteca. Isto, para que sejam traçadas as competências dos profissionais, que além de conhecimentos teóricos e habilidades técnicas, devem ter novas atitudes diante da mudança cultural.

Bawden; Vilar; Zabukovec (2004, p.190) observam que as bibliotecas estão se tornando cada vez mais digitais, e por isso é necessário repensar a idéia de biblioteca como um lugar. Se a idéia de lugar implica localização física, como ela deve ser agora? Lembram que a biblioteca como um espaço de quietude e reflexão deve ser revisto como um ambiente que estimule a criatividade, palavra de ordem no cenário do novo mercado da informação, onde se inserem as bibliotecas digitais. Esse cenário está exigindo uma nova abordagem dos programas de formação para lidar com a informação, que, de acordo com Urs (2002), devem contemplar os seguintes aspectos:

- uma perspectiva interdisciplinar;
- estratégias centradas no usuário (leiam-se estratégias centradas no mercado);
- a filosofia: "informação como recurso";
- foco no conteúdo, independentemente do formato;
- visão do processo de agregação de valor.

Para traçar os programas de formação, Urs (2002) afirma que é fundamental delinear as áreas essenciais (*core areas*) no campo das bibliotecas digitais que ele considera: 1- Usuário da informação; 2- Recursos informacionais; 3- Processos de agregação de valor; 4- Tecnologias da informação; 5- Gestão da informação. Assim, as competências do profissional da informação devem estar diretamente relacionadas com os conhecimentos e habilidades classificados dentro de cada área, como demonstra o quadro 1.

O quadro, apresentado para o manejo de bibliotecas digitais, evoca novamente a idéia de desafio. A complexidade de conhecimentos e habilidades necessários reforça a idéia da importância das atitudes, item que deve compor o conceito de competências necessárias: conhecimentos, habilidades e atitudes. A flexibilidade será altamente desejável, visto que invariavelmente é necessário o trabalho de equipes multidisciplinares.

Quadro 4 – Tarefas e conhecimentos necessários das cinco áreas essenciais (*core areas*)

Area	Tarefas	Conhecimento e Habilidades Requeridos
Usuário da informação	- Conhecer o usuário - Identificar e analisar as necessidades do usuário - Realizar entrevistas de referência - Formular estratégias de busca	- Abordagem teórica referente a estudos de usuários - Métodos de pesquisa qualitativos - Métodos de pesquisa quantitativos - Ferramental da psicologia - Estratégias de busca
Recursos informa-Cionais	- Identificação - Avaliação - Seleção	- Impressos e eletrônicos - Natureza e características - Meta-recursos - Busca e mecanismos de busca - Normas, Técnicas e procedimentos de avaliação
Processos de agregação de valor	- Compilação de bibliografias, bases de dados - Desenvolvimento de bibliotecas virtuais - Design e desenvolvimento de Interfaces para usuários - Organização de recursos informacionais (classificação; catalogação; indexação, elaboração de resumos) - Análise e consolidação do informação - Design e desenvolvimento de conteúdos eletrônicos - Desenvolvimento de bibliotecas digitais	- Fundamentos teóricos da organização do conhecimento (lógico, lingüística, psicologia) - Paradigmas da organização do conhecimento - Linguagens de indexação, ferramentas e técnicas de representação do conhecimento; - Recursos de catalogação e descrição: formatos de registro, metadados, organização de recursos no Internet - Linguagens de programação, DBMS - Ferramentas de autoria; Ferramentas de multimídia e hipermídia - Publicação eletrônica, SGML, HTML, DHTML, XML, PDF - Técnicas de digitalização
Tecnologias da informação	- Estudo de sistemas - Tecnologia da informação para levantamentos de mercado - Identificação e seleção de hardware e software amigáveis para automação, LANs, Internet, Intranets, Extranets, etc. - Ajustes do sistema - Conversão de dados - Desenvolvimento e registro de web sites	- Compreensão de Tecnologia de informação e comunicação. Hardware: computação, telecomunicações, redes e tecnologias de armazenamento - Roteadores/processadores de redes, gateways, bridges, etc., - Sistemas operacionais: Windows, Unix, Linux - Sistemas de operação de redes - Automação de escritórios - Topologia de redes e protocolos - OOPS, C, C++, Visual Basic, Java, ASP, API, PHP, etc - CGI scripting languages: PERL, Java scripting, - Gestão de web site - Geração e validação de documentos digitais - Técnicas de preservação – digital watermarking and digital signatures
Gestão da Informação	- Planejamento estratégico & tático - Gestão de recursos financeiros e humanos - Desenvolvimento de sistemas de suporte à decisão - Reengenharia organizacional - Gestão do conhecimento; gestão do fluxo da informação - Workflow Gestão do fluxo do trabalho - Gestão de projetos - Pesquisa de mercado - Preço e marketing - Avaliação - Certificação de qualidade	- Teorias e escolas de administração - Economia da informação - Psicologia de consumidores - Princípios e técnicas de gestão - Métodos de previsão - Princípios de marketing - Técnicas e ferramentas de pesquisa de mercado - Estratégias e métodos de marketing - Ferramentas de gestão de projetos - Técnicas de orçamentos: técnicas de OR, programação linear, - Avaliação de performance: técnicas de medida - Métodos de qualidade total: série ISO 9000 e outros padrões - Direitos de propriedade e intelectual

Fonte: Tradução livre de: URS (2002, p.2-3)

A visão sistêmica e o comportamento pró-ativo também são fatores fundamentais.

Quanto à adoção de tecnologias não parece ser um complicador sério para os profissionais da informação, relembrando a colocação de Bawden; Vilar; Zabukovec (2004), para quem os profissionais da informação sempre procuram usar as tecnologias disponíveis "em seus tempos". O que significa dizer que a área da informação sempre buscou a tecnologia mais atualizada para organizar, armazenar e recuperar a informação. Deve-se salientar que sempre houve, na verdade, uma preocupação maior com a organização dos estoques informacionais, procurando resolver mais os problemas operacionais do que as demandas dos usuários. Os autores colocam que, dessa perspectiva, as novas tecnologias são como as anteriores, porém o enfoque se redireciona com a mudança, já comentada, do mercado da informação, e também porque mudam as demandas, o que conseqüentemente deve refletir na educação e nas competências dos profissionais da informação.

Nessa perspectiva, Cloonan; Dove (2005, p.1) destacam que este é um bom tempo para se retomar a ideologia clássica das cinco Leis da Biblioteconomia, propagada por Ranganathan: 1- livros são para serem usados; 2- a cada leitor o seu livro; 3- a cada livro o seu leitor; 4- poupe o tempo do leitor; 5- a biblioteca é um organismo em crescimento. Eles destacam que as cinco leis continuam em perfeita consonância com esse novo meio digital, para o qual devem adaptar-se os profissionais da informação. Enfatizam a consideração da 3ª lei porque ela tem particular relevância na crescente proliferação de recursos no meio digital, já que mecanismos de recuperação da informação como o Google e o Yahoo são cada vez mais populares e, a princípio, levam o usuário a encontrar de forma fácil o que necessitam, ou seja, dispensam a intermediação do profissional.

No entanto, a informação relevante não é tão facilmente recuperada pelo usuário. Os autores enfatizam que é possível, e mesmo necessário, aplicar a 3ª lei no meio digital, ou seja, fazer com que os recursos informacionais estejam de tal forma bem organizados que "eles encontrem os seus usuários", num paralelo a cada livro o seu leitor. Para isso propõem a consideração de cinco aspectos, que são comentados a seguir:

1 -**Conexão máxima**: construir tantas possibilidades de acesso quanto possíveis, ou seja, maximizar a possibilidade de localização ao usuário

para o que ele necessita. Essa é uma importante tarefa do profissional nesse meio, cujo maior desafio continua sendo a recuperação da informação relevante. Dessa forma, Cloonan; Dove (2005) apontam que a 3ª lei pode ser subvertida para: "os recursos informacionais encontram seus usuários". No meio digital isso é possível com a colocação de *links* que levam a recursos similares àqueles que os usuários buscam, a exemplo do serviço oferecido pela Amazon.com, que traça comentários e leva a outros itens que tratam do assunto procurado.

2-Construir uma "Coleção eletrônica bem arranjada: significa aplicar a 3ª. lei ao princípio da valorização do conteúdo, de forma rápida e fácil para o usuário, concorrendo com os mecanismos de busca que trazem uma avalanche de informação. No entanto, ao contrário de mecanismos como o Google, a construção de uma "estrutura de caminhos" deve ser por esquemas bem elaborados, que contribuam para uma recuperação relevante. Nos mecanismos de busca disponíveis na Internet, a análise e a indexação da informação são operações automáticas, realizadas por robôs. Esses robôs utilizam a técnica de extração de palavras para representar o conteúdo do texto analisado. Ela é diferente da técnica dos indexadores humanos, que é a da atribuição, que representa o conteúdo com termos significativos, mesmo que esses termos não apareçam no objeto analisado, assim como outros aspectos que não estão explícitos, como por exemplo: gráficos; figuras; tipo de contexto; tipo de *site* e outros. Outra limitação dos indexadores automáticos é que reconhecem somente o texto. A riqueza de informações contidas em imagens, gráficos e vídeo ficam de fora. Alguns programas podem encontrar cores e padrões de imagem, mas nenhum pode fazer deduções e relações de significado de uma imagem. Isso é ainda domínio somente dos humanos, pois depende de senso crítico, *feeling*. Essas capacidades ainda não foram incorporadas aos *softwares* (Lynch, 1997, apud Silva, 2000).

3-A biblioteca invisível: a expressão *Web Invisible* foi cunhada por Sherman; Price (2001, *apud* Cloonan; Dove, 2005, p.3), alertando para o fato de que os mecanismos de busca deixam grande parte da Internet sem indexação. Ficam de fora fontes de informação valiosas, como as governamentais, universitárias e de mercados menos visíveis. Algumas iniciativas fazem tentativas no sentido de levar outros recursos não cobertos por esses mecanismos aos usuários, a exemplo do Google Print e

do Open WorldCat da OCLC (*On-line Computer Library Center*). A grande questão é que esses recursos não são conectados, ou integrados por "caminhos de mão dupla". Também em muitas bibliotecas digitais isso acontece. Existe dentro delas "salas fechadas", onde conteúdos ficam sem acesso, a exemplo das bibliotecas com paredes. Os conteúdos precisam estar disponíveis num fluxo contínuo, livre. Para isso é importante a adoção de padrões de metadados que permitem a integração de vários sistemas, ou bibliotecas digitais, ou recursos informacionais, a exemplo do formato MARC.

4-Navegação eletrônica (*electronic browsing*): os profissionais devem aproveitar as possibilidades do *hiperlink*. A navegação em rede permite que o leitor vá de recurso em recurso de informação, fazendo com que ele chegue no que sabe que quer e naquilo que não conhece e atende às suas necessidades. Assim, os profissionais devem estender seu trabalho no sentido de ligar o maior número possível de peças de informação. Os recursos podem ser bem organizados e integrados, facilitando a navegação do leitor. Assim é possível ligar autores, editores, livreiros, pesquisadores etc. A facilidade de acesso e de forma rápida é que os usuários das redes esperam. (Cloonan; Dove, 2005, p.4)

5-A Web semântica: o conceito de Web semântica está estreitamente relacionado à questão principal da avalanche informacional, possível de ser disponibilizada nos sistemas automatizados e ligados em rede. Essa idéia prende-se à faculdade humana de estabelecer significados ou captá-los onde não estão explicitamente colocados. Voltamos ao exemplo da indexação: os mecanismos indexadores falham na sua tarefa porque não possuem essa faculdade da abstração e do *feeling*, como já colocado. A informação disponível na Internet não está colocada dentro de estruturas semânticas. Com esse objetivo foi criada a idéia da *Web* semântica, que visa a definir o significado da informação através de ontologias – documento que descreve um vocabulário de termos para comunicação entre humanos e agentes automatizados.

A idéia disseminou-se principalmente depois do artigo de Berners-Lee, Hendler, Lassila, "*The Semantic Web*", publicado em 2001, na revista *Scientific American*. Berners-Lee é o inventor da *World Wide Web* e atualmente um dos diretores da W3C, organização para o desenvolvimento da *Web*, que promove uma das principais iniciativas para estudo e desenvol-

vimento da *Web* semântica. Bax (2004) coloca que os autores acima definem a Web semântica como uma extensão da Web tradicional, onde, a partir do uso intensivo de metadados, espera-se obter o acesso automatizado às informações, com base no processamento semântico de dados e heurísticas feitos por máquinas. Para tal, desde 1998, a equipe da W3C (World Wide Web Consortium) vem trabalhando arduamente no desenvolvimento de tecnologias avançadas, que visam à representação estrutural e semântica dos recursos na Web. Essas tecnologias, aliadas à teoria de domínios ou ontologias, permitem oferecer um serviço com um nível maior de qualidade. Dentro destas perspectivas, a Web será capaz de tecer uma rede extensa de conhecimento humano, podendo ainda, por meio do processamento via máquina, inferir novos conhecimentos.

Para Levy (apud Bax, 2004) a Web semântica como "ferramenta" é necessária para o desenvolvimento da inteligência coletiva. Segundo ele, a semântica da Web é a elaboração de um sistema de códigos. Não é necessário que as pessoas utilizem uma nova língua – cada um utilizará a sua – , mas haverá um *software* que traduzirá o que está sendo escrito ou dito para uma língua universal. Ranganathan afirmava que o negócio do bibliotecário é adotar todos os métodos possíveis para atrair seu público real e potencial. No meio digital isso é possível com esse serviço de valor agregado e customizado que amplia as chances do usuário. A 5ª lei tem tudo a ver com a 3ª: oferecer serviços é a razão de bibliotecas e bibliotecários.

Quadro 2 – Os Imperativos da Cinco Leis – Antes e Agora

As Leis	Na época de Ranganathan	Atualmente
Livros são para serem usados	Colocar os livros em circulação não significa o mesmo que compartilhar	Os recursos *on-line* estão disponíveis onde e quando forem necessários
A cada leitor o seu livro	Quebra as barreiras do princípio da educação (e livros) para todos	Eliminar os obstáculos que impedem os usuários de fazerem uso efetivo dos recursos eletrônicos
Para cada livro o seu leitor	Abrir os depósitos e preparar catálogos cruzados eficientes	Integrar recursos eletrônicos num ambiente virtual de aprendizagem
Poupe o tempo do leitor	Criar catálogos efetivos para a localização rápida de um livro em particular	Prover metabuscadores capazes de oferecer ao usuário a possibilidade de pesquisar em todos os recursos disponíveis e links identificando as melhores fontes
A biblioteca é um organismo em crescimento	Ver as bibliotecas como parte de uma comunidade maior	Oferecer serviços 24 horas, acessíveis de qualquer lugar, promover o acesso a Web invisível

Fonte: Cloonan; Dove (2005,p.5)

Os autores sintetizam a aplicação da 3ª. lei no meio digital com as seguintes proposições:

- implementar tantos caminhos quantos você possa para leitores encontrarem qualquer recurso;
- foco em obter recurso que permita um *link* semântico, não somente no nível do título mas da palavra e da frase;
- estabelecer portais de assuntos específicos;
- divulgar as partes mais importantes dos recursos informacionais pelas páginas ou *e-mail*;
- engajar professores e *experts* no assunto trabalhado, para estabelecer links na biblioteca digital;
- prover as pessoas de serviços de suporte de referência a qualquer momento, via *e-mail* ou *chats*;
- colocar toda a atenção no mercado: fazer tudo que comerciantes fazem e mais ainda para capturar o interesse daqueles que poderiam se beneficiar muito de seus recursos eletrônicos, mas podem não conhecer esses recursos ou não saber usá-los.
- "alcance" seus usuários onde estiverem, via *on-line*!

4. Considerações finais

Parece possível colocar um marco de compreensão para a relação biblioteca digital e profissional da informação, tomando como base a teoria das 'redes complexas'. Do ponto de vista dessa teoria:

> Muitas estruturas ou fenômenos naturais – como nossos pensamentos, nossos relacionamentos, diferentes sistemas biológicos, o conjunto das estradas de um país – têm a forma de redes de múltiplas conexões, conhecidas como redes complexas. A pesquisa teórica nessa área vem apontando modelos e propriedades dessas redes, que permitem simular fenômenos naturais e estruturas criadas pelo homem, ajudando a entender melhor diversos aspectos de nossa vida e do mundo que nos cerca. (Costa, 2005, p. 34)

Assim, o conceito de biblioteca digital pode se fixar como a organização da informação de forma customizada e numa complexa rede, ou sub-rede,

que interconecta todos os recursos de informação. Nesse sentido, pode-se considerar as competências do profissional da informação no papel de interconector, estabelecendo nós (*hubs*) de forma a conseguir a máxima conexão, anunciada por Cloonan; Dove (2005), na observância da 3ª lei de Ranganathan. Na verdade, como observam os autores, essa ação subverte a 3ª lei, no sentido de que os recursos vão ao encontro dos seus usuários. É como se, à medida que o usuário vai entrando, os recursos "pulassem" ao seu encontro. É o dar visibilidade à Web invisível, como também propõem os autores. É um trabalho de aranha: o tecer infinito de recursos informacionais e ou interações sociais. A complexidade está, como afirma Costa (2005, p.36), no grande número de nós e conexões existentes e possíveis.

Estes são novos desafios que se apresentam de forma tão diferente daqueles que se apresentavam à práxis bibliotecária anterior, amparada num saber consolidado, até que as mudanças se acelerassem. Desafios trazidos pela nova cultura e que poderão ser transpostos pela adoção de novas posturas profissionais que estão a exigir novas competências conceituais para os trabalhos de análise e resolução de problemas:

- de avaliação dos serviços, do conteúdo e do valor da informação;
- de avaliação das necessidades informacionais do ambiente interno e externo às organizações;
- de criação de métodos e indicadores para avaliação da qualidade dos serviços informacionais;
- de criação de pacotes informacionais dirigidos para públicos distintos etc.

Exigem também competências técnicas, que se expressam nas respostas dadas às tarefas de processamento e gerenciamento da informação:

- de custos, ligados ao acesso à informação *on-line*, que não se apresenta de maneira gratuita, como se possa pensar. Há um custo de investimento na compra de equipamentos, na aquisição de assinaturas de materiais eletrônicos, no acesso às redes, o que se reflete sobre as novas maneiras de traçar as políticas de aquisição e de desenvolvimento de coleções;
- para resolver dificuldades na preservação da informação, pois não há garantia de que o documento acessado num tempo permaneça o

mesmo em outro momento;

- de garantia da integridade do documento, sua fluidez (Levacov, 1997).

Mas, não só! São extremamente importantes, para a instauração de uma nova ética social, as competências interpessoais, relacionais, comunicacionais, que se referem à capacidade de cooperar e trabalhar em equipe e de conviver com os outros. As competências relacionais se expressam pelas atitudes que favoreçam o melhor relacionamento entre pessoas, gerando o espírito de compartilhamento não só de tarefas técnicas mas, principalmente, engendrando melhores ações no convívio entre as pessoas, o que é necessário para a continuidade da existência em sociedade.

Nota

[1] http://www.scielo.br/cgi-bin/wxis.exe/iah/

Referências

ARRUDA, Maria da Conceição Calmon; MARTELETO, Regina Maria; SOUZA, Donaldo Bello. Educação, trabalho e o delineamento de novos perfis profissionais: o bibliotecário em questão. *Ciência da Informação*, Brasília, v. 9, n. 3, p.14-24, set./dez. 2000.

BAX, Marcello. *Segundo Tim Berners-Lee, a Web semântica é uma extensão da Web tradicional*. Disponível em: <http://www.bax.com.br/news/News_Item.2004-04-29.8261853316> Acesso em: 28 abr. 2005.

BAWDEN, David; VILAR, Polona; ZABUKOVEC, Vlasta. Competences and capabilities for the digital library. On-line Information, 2004. *Proceedings*. Disponível em: www.online-information.co.uk/2004proceedings/ thursam/ bawden_vilar_zabukovec.pdf > Acesso em: 28 abr. 2005.

Berners-Lee, Tim; Hendler, James; Lassila, Ora. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, May, 2001. Disponível em: <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2> Acesso em: 28 abr. 2005.

CANIELLO, M. O ethos sanjoanense: tradição e mudança em uma cidade pequena. *Mana*: estudo de antropologia social. v.9, n.1, p.31-56, 2003

CAPURRO, Rafael. Perspectivas de una cultura digital en Latinoamerica. *Datagramazero*-Revista de Ciência da Informação, v.3, n.2, abr. 2002. Disponível em: <http://www.dgz.org.br/abr02/F_I_art.htm> Acesso em 30 mar. 2005.

CLOONAN, Michele V.; DOVE, John G. Ranganathan online: do digital libraries violate the Tird Law. *Library Journal*, April 2005. Disponível em: <http://www.libraryjournal.com/article/CA512179> Acesso em: 2 abr. 2005.

COSTA, Luciano da Fontoura. Redes complexas: modelagem simples da natureza. *Ciência Hoje*, v.36, n.213, mar. 2005. p. 34-39.

CUNHA, Murilo Bastos. Desafios na construção de uma biblioteca digital. *Ciência da Informação*, Brasília, v. 28, n. 3, set./dez. 1999. p. 257-268.

FERREIRA, F. et al. Política de formação de profissionais e competências essenciais. In: Cinform: Encontro Nacional de Ciência da Informação, 5, Salvador, 2004. *Trabalhos*.

Disponível em: <http://www.cinform.ufba.br/v_anais/frames.html> Acesso em: 12 abr. 2005.

LASTORIA, Luiz. A. C. Nabuco. Ethos sem ética: a perspectiva crítica de T.W. Adorno e M. Horkheimer. *Educação e Sociedade*, v. 22, n. 76, out. 2001.

LEVACOV, Marília. *Bibliotecas Virtuais: ® evolução?. Ciência da Informação*, Brasília, v.26, n.2, maio/ago.1997.

MACIAS-CHAPULA, César A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. *Ciência da Informação*, Brasília, v. 27, n. 2, maio/ago, 1998. p.134-140.

MARTINS, R. D. Perspectivas para uma biblioteca no futuro: utopia ou realidade. *Revista Informação & Sociedade: Estudos*.João Pessoa, v.12, n.1, 2002. Disponível em: <http://www.informacaoesociedade.ufpb.br/1210206.html> Acesso em 30 abr. 2005.

RAMOS, Paul A. Baltazar A gestão na organização de unidades de informação. *Ciência da Informação*, Brasília, v.25, n.1, jan./abr.,1996. p.15-25.

SILVA, H.P. *Inteligência competitiva na Internet;* proposta de um processo. 2000. 185p. Tese (Doutorado em Engenharia de Produção) - Departamento de Engenharia de Produção e Sistemas, Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis.

SILVA, Neusa; SÁ, Nysia O.; FURTADO, Sandra R.S. Bibliotecas digitais: do conceito às práticas. In: Simpósio Internacional de Bibliotecas Digitais, Campinas, 2004. *Trabalhos apresentados*..Disponível em: <http://libdigi.unicamp.br/document/?code=8304> Acesso em: 25 abr. 2005

TAYLOR, Arlene G. *The organization of information*. 2.ed. Westport: Libraries Unlimited, 2004.

THOMPSON, John B. *A mídia e a modernidade: uma teoria social da mídia*. 2.ed. Petrópolis, Vozes, 1998.

VARGAS ZÚÑIGA, Fernando. De las virtudes laborales a las competencias clave: un nuevo concepto para antiguas demandas. *Boletin Cinterfor*, n. 149, p. 9-23, may./ago., 2000. Disponível em: http://www.cinterfor.org.uy/public/spanish/region/ampro/cinterfor/publ/boletin/149/pdf/f_varg.pdf> Acesso em: 20 maio 2004.

VARGAS ZÚÑIGA, Fernando. *La formación por competências. Instrumento para incrementar la empleabilidad*, 2001. Disponível em: http://www.cinterfor.org.uy/public/spanish/region/ampro/cinterfor/publ/sala/vargas/for_comp/ Acesso em: 20 maio 2004.

URS, Shalini R. Redefining, reinventing and repositioning the information professionals and digital libraries in the new information landscape. In: WORKSHOP ON DEVELOPING DIGITAL LIBRARIES, EDUCATION AND TRAINING PROGRAMS, JOINT CONFERENCE ON DIGITAL LIBRARIES,

2002. Portland. *Proceedings of the...*,2002. Disponível em <http://lair.indiana.edu/papers/urs.doc> Acesso em: 30 mar. 2005.

WATSTEIN, Sarah B.; CALARCO, Pascal V.; GHAPHERY, James S. Digital library: keywords. *Reference Services Review*, v.27, n.4, 1999.

5. Experiências brasileiras e internacionais

Biblioteca digital brasileira: integrando a ICT brasileira

Hélio Kuramoto *

1. Introdução

O surgimento das tecnologias da informação e da comunicação, dentre as quais aquelas que suportam a rede Internet, possibilitaram o registro e a disseminação de um volume cada vez mais crescente de informação, em suas mais variadas formas. Esse fenômeno não é isolado, ele acontece em todos os cantos do globo terrestre. No entanto, a participação brasileira em *sites web*, em 1999, segundo Edward T. O'neill, Brian F. Lavoie; Rick Benett [4], era de 2%, enquanto os EUA detinham 49% dos *sites web* públicos. Ainda, segundo esses mesmos autores, em 2002 a participação dos EUA era de 55%, enquanto a participação brasileira não mais aparecia na mesma estatística (Gráfico 1). Em termos de conteúdos em língua portuguesa, em 1999, segundo esses mesmos autores, haviam cerca de 2%, enquanto em 2002 essa porcentagem caia para 1%, mantendo-se a língua inglesa com os mesmos 72%, tanto em 1999 quanto em 2002.

As tecnologias da informação e da comunicação facilitaram sobremaneira os procedimentos de tratamento, registro e disseminação da informação por meio da rede Internet. Além das facilidades proporcionadas por essas tecnologias, estas se tornaram mais baratas e populares. Hoje

*Doutor em ciências da informação e da comunicação pela Université Lumière – Lyon 2, França. Coordenador geral de projetos especiais e do Programa Biblioteca Digital Brasileira, do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). kuramoto@ibict.br

não é mais necessário muito investimento para se colocar uma base de dados para consulta em linha, via Internet. Deve-se, no entanto, ressaltar que esta é uma realidade para as instituições localizadas em países desenvolvidos, o mesmo ainda não acontece em países como o Brasil. Infelizmente as bibliotecas e centros de documentação das instituições de ensino superior ou mesmo de institutos de pesquisa ou de outros órgãos governamentais não contam com a mesma facilidade devido aos parcos orçamentos de que se dispõem para o tratamento, orgnaização e disseminação da informação. E muitas vezes, quando o fazem, muitas não adotam padrões e protocolos que possibilitem a interoperabilidade de seus recursos de informação com outros. Nesse contexto, deve-se ressaltar que muitas das bibliotecas universitárias automatizaram os seus acervos, sem, entretanto, adquirir ou implantar o protocolo Z39.50, o qual permitiria manter os respectivos acervos interoperáveis com outros sistemas de informação. Existe, portanto, uma heterogeneidade de soluções tecnológicas, a qual vem dificultando o acesso à informação.

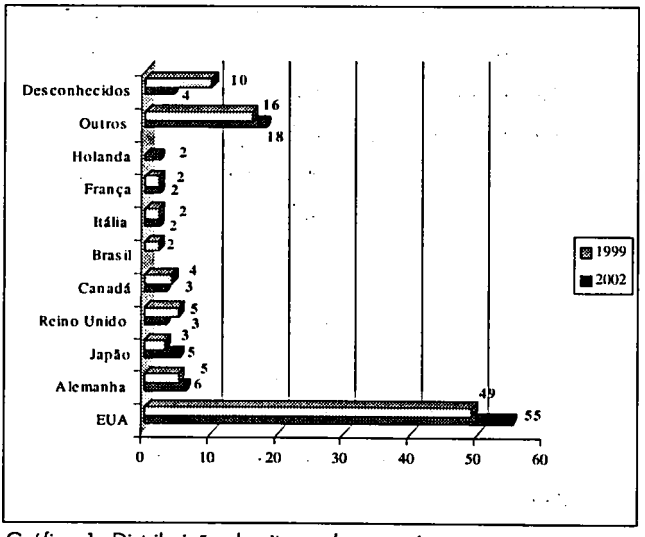


Gráfico 1. Distribuição de *sites web* por países

Mais que as facilidades tecnológicas e as dificuldades orçamentárias, vale lembrar que o registro, o tratamento e a disseminação da informação científica e tecnológica são fundamentais para o desenvolvimento científico e tecnológico do país.

Pelo cenário apresentado, conclui-se que há necessidade de iniciativas que estimulem o incremento do registro de conteúdos brasileiros na rede Internet e, por conseguinte, em língua portuguesa; que facilite o acesso à informação à comunidade científica e tecnológica, em particular, e à sociedade em geral; que se assimilem as novas tecnologias da informação e da comunicação para o tratamento, organização, registro e disseminação da informação científica e tecnológica, colaborando para a expansão da presença brasileira na Internet e, principalmente, para o desenvolvimento científico e tecnológico do Brasil. É com esse propósito que o Instituto Brasileiro de Informação em Ciência (IBICT) vem desenvolvendo esforços para a construção da Biblioteca Digital Brasileira.

2. Contexto conceitual

O projeto Biblioteca Digital Brasileira foi concebido com base no conceito de que uma biblioteca digital é um sistema de informação que envolve coleções compostas tanto por informações referenciais (metadados) e conteúdos integrais, quanto apenas por informações referenciais, com o propósito de facilitar o acesso à informação à comunidade científica e tecnológica brasileira. Essa concepção foi adotada com o propósito de integrar também as OPACs (On-line Public Access Catalog) das bibliotecas universitárias brasileiras com vistas à futura implantação de um serviço de empréstimo entre bibliotecas via o programa Comut.

3. Objetivos

Após a introdução apresentada nesse documento, talvez fosse dispensável explicitar novamente os objetivos do projeto da Biblioteca Digital Brasileira. Entretanto, por uma questão de clareza, os objetivos serão enumerados a seguir:

- aumentar a presença brasileira na Internet, por meio do incremento de registro e disseminação de documentos eletrônicos (textos completos);
- criar mecanismos que possibilitem, aos usuários, acessar recursos informacionais distribuídos e heterogêneos por meio de uma interface única;

- integrar recursos informacionais heterogêneos;
- tratar, organizar, registrar e disseminar a produção científica e tecnológica brasileira.

Esses objetivos sintetizam a preocupação do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) com a informação científica e tecnológica. Ao mesmo tempo em que se preocupa com o aumento da presença brasileira na Internet, o Instituto busca cumprir com a sua missão, que é promover e estimular o registro e a disseminação da produção científica e tecnológica brasileira.

4. Biblioteca digital brasileira: subprojetos

Para atingir os objetivos estabelecidos para a BDB, foram definidos quatro subprojetos, os quais são descritos a seguir:

- Biblioteca Digital de Teses e Dissertações (BDTD)

A BDTD se propõe a registrar, tratar, organizar e a disseminar parte da produção científica brasileira, as teses e dissertações.

- Diálogo Científico

Essa iniciativa tem o propósito de inovar os procedimentos da comunicação científica, registrando-a e dissminando-a. O Diálogo Científico é a recriação do colégio invisível por meio do uso das tecnologias da informação e da comunicação.

- Catálogo de Anais de Congresso Eletrônico

Subprojeto concebido em parceria com o Centro de Informações Nucleares da Comissão Nacional de Energia Nuclear. A idéia dessa iniciativa é o desenvolvimento de um pacote de *software* que seja capaz de criar, gerir e publicar anais de congresso em meio magnético, disponibilizando-os em uma biblioteca digital.

- Integração de recursos de informação

Hoje a grande dificuldade encontrada pelos usuários de informação é a desorganização causada pelo crescente volume de informações que é colocado disponível na Internet. Para se fazer a mesma busca em várias bases de dados, o usuário tem de fazer acesso a cada um dos *sites* onde

se encontram os vários acervos de informação. A idéia dessa iniciativa é o desenvolvimento de uma única interface de busca que possibilite ao usuário submeter uma mesma busca a várias bases de dados distribuídas pela rede Internet. Portanto, o desenvolvimento dessa interface permitirá a integração dos vários recursos de informação existentes no país, independentemente das tecnologias utilizadas pelos provedores a serem integrados.

Mais recentemente, o IBICT vem trabalhando em um outro projeto a ser integrado à BDB. Trata-se de um servidor de publicações periódicas eletrônicas. O objetivo dessa iniciativa é colaborar com os editores científicos na manutenção de suas publicações periódicas, oferecendo-lhes um ambiente no qual ele possa publicar e gerir todo o processo de submissão, seleção e publicação dos artigos. Para tanto, o IBICT customizou o pacote de *software Open Journal Systems*, o qual ganhou a denominação, em português do Brasil, de Sistema Eletrônico de Editoração de Revistas (SEER).

5. Princípios

O IBICT adotou alguns princípios que fundamentam todo o processo de desenvolvimento e implantação da BDB, os quais seguem:

Desenvolvimento e uso preferencial de *software open source*

A experiência tem mostrado que as instituições mantedoras de acervos de informação encontram dificuldades para disponibilizar e disseminar os seus acervos utilizando as novas tecnologias da informação e da comunicação. Além disso, essas instituições não contam com quadros de pessoal especializado em tecnologias da informação para o desenvolvimento das ferramentas necessárias. A partir dessa constatação, o IBICT adotou esse princípio com o propósito de transferir os pacotes de *software* para que essas instituições possam implantar as aplicações de bibliotecas digitais, bases de dados etc. A utilização de ferramentas de *software* proprietário dificultaria o repasse de pacotes de *software* para outras instituições. A adoção de *software open source* possibilitou, por exemplo, a distribuição do pacote de *software* TEDE, pacote para publicação eletrônica de teses e dissertações.

Adoção de modelo distribuído

No mundo de hoje, com as tecnologias da informação disponíveis, não se pode admitir modelos centralizados para sistemas de informação. Além disso, o Brasil tem dimensões continentais e os acervos de informação encontram-se em várias instituições, portanto a adoção de um modelo distribuído com tecnologias que possibilitem integrar esses acervos facilita todo o processo de sensibilização e valorização dessas instituições. Nesse processo, o IBICT passa a ter o papel de facilitador e integrador.

Adoção de padrões e protocolos de ampla aceitação internacional

Trata-se do principal fundamento para a obtenção de sucesso em uma iniciativa como o da Biblioteca Digital Brasileira. A adoção de padrões e protocolos de ampla aceitação internacional, como o XML, o *Dublin Core*, permite obter maior interoperabilidade entre os vários sistemas de informação.

6. Metodologia

O projeto da Biblioteca Digital Brasileira está estabelecido com base em duas vertentes principais: 1) a integração de fontes heterogêneas de informação; 2) a construção de repositórios para abrigar e disseminar conteúdos científicos e tecnológicos brasileiros na Internet.

A integração de fontes heterogêneas de informação propiciará o desenvolvimento de uma interface única para acesso às várias fontes de informação existentes no país e no exterior.

Existem basicamente três protocolos que permitem a busca integrada a diversas fontes de informação: 1) Z39.50; 2) HTTP; 3) OAI-PMH.

O protocolo Z39.50 é utilizado normalmente pelas OPAC (*On-line Public Access Catalog*). A maioria dos pacotes de *software* para automação de bibliotecas utiliza esse protocolo. Portanto, a maioria das OPAC pode ser integrada por meio desse protocolo.

O protocolo HTTP (*Hypertext Transfer Protocol*) é utilizado em toda a *Web* para possibilitar a navegação hipertextual nas páginas ou *sites web*. Portanto, todas as fontes de informação que se encontram na *Web* utilizam esse protocolo.

O OAI-PMH (Open Archives Initiative – *Protocol of Metadata Harvest*) é um protocolo utilizado pelos repositórios que utilizam os padrões da iniciativa de arquivos abertos. Esse protocolo é utilizado para expor e capturar metadados dos repositórios que utilizam o padrão e os ideais dos arquivos abertos. Esse protocolo funciona de forma aliada ao HTTP.

Considerando a existência de sistemas que oferecem essa facilidade de integração de fontes de informação, os chamados metabuscadores, dentre os quais destacam-se o QueryServer e o Metalib. Por um processo licitatório, a solução que saiu vencedora foi o Metalib. Esse pacote de *software* é utilizado por diversas universidades estrangeiras. Trata-se, é bem verdade, de uma solução proprietária. Até o presente momento, não se identificou qualquer solução de *software open source* que atendesse às características e funcionalidades oferecidas pelo Metalib.

Para a segunda vertente, a criação de repositórios, foram adotados os preceitos e padrões do *Open Archvies Initiative* (OAI), tendo em vista o fato de vários projetos de bibliotecas e repositórios digitais o adotarem, inclusive, o bem-sucedido projeto NDLT|D – *Networked Digital Library of Thesis and Dissertation*. Essa biblioteca digital foi desenvolvida pela Virginia Tech e integra, em uma biblioteca digital federada mais de 100 instituições de ensino superior, inclusive o IBICT, por meio da BDTD. Os resultados obtidos pela BDTD têm sido satisfatórios, pois o modelo mostrou que funciona, tendo o Instituto conseguido integrar, por meio de *harvest*, atualmente 12 instituições de ensino superior. Algumas dessas instituições, como o Instituto Nacional de Pesquisas Espaciais (Inpe) e a Unicamp, apesar de utilizarem pacotes de *software* próprios ou diferentes daquele desenvolvido pelo IBICT (TEDE), conseguiram se integrar à BDTD pelo fato de terem implementado o protocolo OAI-PMH e adotado o padrão de metadados para descrição de teses e dissertações (MTD-BR). Verifica-se, portanto, o acerto em ter adotado esse modelo em função do alto nível de interoperabilidade alcançado.

Os padrões adotados e preconizados pela OAI são:

- Dublin Core como padrão de metadados para descrição dos objetos digitais;
- Protocolo OAI-PMH –*Open Archives Initiative Protocol of Metadata Harvest;*

• XML como linguagem de marcação para encapsular os registros dos objetos digitais.

Os ideais estabelecidos pela OAI são:

• modelo distribuído, também chamado de federado, composto por dois atores principais[1]: provedores de dados e provedores de serviços;

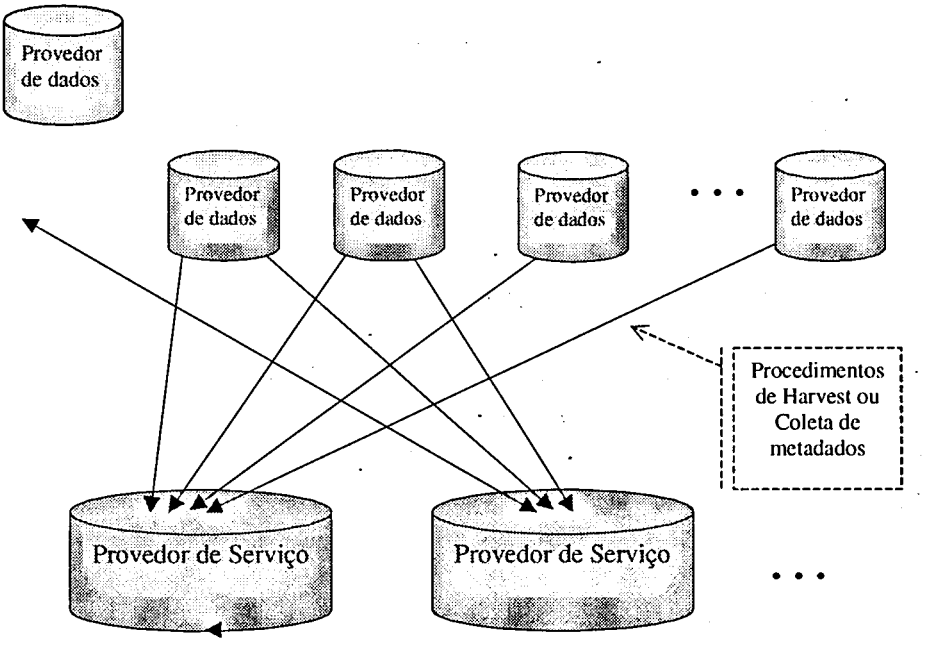
Os provedores de dados são os gestores de arquivos de *e-prints* ou também denominados de repositórios, aos quais são submetidos, diretamente pelos seus autores, os *pre-prints*, relatórios técnicos, dentre outros tipos de documentos. Os provedores de dados tornam disponíveis, à comunidade, mecanismos para submissão ou auto-arquivamento, responsabilizam-se pelo armazenamento dos trabalhos a longo prazo, preocupam-se com a preservação dos objetos digitais e tornam disponíveis, para coleta, os metadados relativos aos trabalhos ou documentos publicados no seu repositório.

Os provedores de serviços, que também podem caracterizar-se como agregadores, são instituições que se preocupam em coletar os metadados armazenados nos vários provedores de dados, oferecendo uma interface ou mecanismo de busca que possibilita aos usuários finais a consulta ou busca em um repositório que integra os metadados coletados a partir de diferentes provedores de dados.

• auto-submissão[2] de documentos eletrônicos;
• compromisso de armazenamento a longo prazo;
• uso e desenvolvimento de *software open source*[3];
• acesso livre à informação;
• uso de padrão de metadados, normalmente baseado no Dublin Core;
• coleta de metadados por meio de um protocolo de comunicação denominado protocolo de *harvesting*, também chamado de OAI-PMH (Open Archives Initiative – Protocol of Metadata Harvesting). Esse protocolo possibilita a interoperabilidade entre os vários provedores de dados e serviços.

O esquema de funcionamento do modelo de *Open Archives* é mostrado na figura 1.

Figura 1 – Esquema de funcionamento dos *Open Archives*



No esquema da figura 1, o leitor pode observar a interoperabilidade entre um provedor de dados e um provedor de serviço, por meio dos procedimentos de *harvest*, também chamado de exposição (por parte dos provedores de dados) e coleta (por parte dos provedores de serviços). Assim, verifica-se que os provedores de serviços podem ser caracterizados como sendo agregadores, tendo em vista que o mesmo integra os metadados de vários provedores de dados. É importante observar que, nesse procedimento de *harvest*, apenas os metadados são coletados, os conteúdos integrais dos documentos são mantidos nos provedores de dados. Portanto, o usuário tem acesso, no provedor de serviço, aos metadados, por meio dos quais ele identifica o documento que atende à sua necessidade de informação e, quando ele tiver interesse em ver o documento em texto integral, ele o verá acionando o *link* para o provedor de dados, que é o local onde estão armazenados os documentos ou objetos digitais.

Uma outra constatação, a partir do esquema da figura 1, é que um provedor de serviço pode desempenhar o papel de provedor de dados em

relação a um outro provedor de serviço, tendo em vista que ele pode expor os seus metadados para esse outro provedor de serviço. Esse papel é o que tem sido desempenhado pelo IBICT no contexto da Biblioteca Digital de Teses e Dissertações. O Instituto é provedor de serviço com relação às instituições de ensino superior, dado que o IBICT coleta os metadados das instituições de ensino superior. Por outro lado, o IBICT desempenha o papel de provedor de dados em relação à NDLTD, dado que ele expõe os seus metadados para a NDLTD coletá-los, inclusive utilizando um outro padrão de metadados, que é o ETD-MS. Observa-se, pela descrição do modelo, a sua versatilidade e potencial de integração.

Diante do exposto, o IBICT, mais do que simplesmente investir no modelo proposto pela *Open Archives Initiative* como um subprojeto da BDB, conforme proposto inicialmente por Trista; Café [7], adotou-o como metodologia para a implementação de respositórios, tanto institucionais quanto temáticos.

7. Estágio atual da biblioteca digital brasileira

Biblioteca Digital de Teses e Dissertações

O IBICT praticamente concluiu o desenvolvimento da Biblioteca Digital de Teses e Dissertações (BDTD), com base no modelo apresentado na metodologia. Antes de iniciar o desenvolvimento das tecnologias necessárias para a implantação da BDTD, o IBICT constituiu um comitê denominado de Comitê Técnico-Consultivo para acompanhar e definir as principais questões técnicas relativas à BDTD. Assim, uma das primeiras tarefas foi a definição do padrão brasileiro para descrição de teses e dissertações, o MTD-BR. Esse padrão foi inicialmente baseado no padrão da NDLTD, o ETD-MS, que por sua vez é baseado no Dublin Core. O padrão brasileiro foi elaborado com vistas à integrar a BDTD a outros sistemas de informação nacional, como a Plataforma Lattes, o sistema da Capes e o SCIELO. Em seguida, o Instituto desenvolveu os seguintes pacotes de *software*:

TEDE (versão descentralizada) – pacote para publicação eletrônica de teses e dissertações

Essa versão implementa a biblioteca digital de teses e dissertações nas instituições de ensino superior (IES), doravante denominado de biblio-

teca digital de teses e dissertações local. Essa foi a primeira versão desse pacote de *software* e contempla a participação de três atores principais: o aluno, a secretaria de pós-graduação e a biblioteca da universidade. O processo de publicação de teses e dissertações se inicia quando o aluno solicita o agendamento da defesa da sua tese ou dissertação. Nesse momento, a secretaria de pós-graduação atribui uma senha ao aluno, que inicia a inserção dos dados cadastrais, conforme o MTD-BR. Uma vez defendida a tese ou dissertação, o aluno faz a sua publicação por meio do *upload* dos arquivos que compõem a sua tese ou dissertação. A partir desse momento, a secretaria de pós-graduação entra em ação para verificar a conformidade dos dados do aluno e do documento (tese ou dissertação), liberando-os para a biblioteca que os complementa com informações específicas de biblioteca. Assim, uma vez que a biblioteca libera o registro da tese ou dissertação, os metadados desta tese ou dissertação tornam-se disponíveis para serem coletados pelo procedimento de *harvest* executado pela BDTD, no IBICT, assim como para consulta, em nível local, e posteriormente em nível nacional, após o *harvest*. O mesmo acontece, em nível internacional, pela NDLTD.

Verificou-se, ao longo dos procedimentos de transferência de tecnologia para as IES, capacitação dos seus técnicos e da implantação, uma certa dificuldade. Isso ficou comprovado ao se observar que, após o treinamento de 162 técnicos de 72 IES, apenas nove deles chegaram efetivamente a implantar esse pacote em suas universidades.

A conclusão, após avaliação desse processo de repasse de tecnologias e implantação nas IES do modelo e da metodologia concebida pela BDTD, é que existe uma cultura nas IES e que a metodologia proposta pela BDTD provoca mudanças de paradigma, pois o próprio autor é quem se responsabiliza por publicar a sua tese ou dissertação. Além disso, a existência de três atores envolvidos nos procedimentos de publicação das teses e dissertações exige maior interação entre esses atores, além de padrões, procedimentos de controle e acompanhamento.

O pacote de *software* TEDE integra, além dos mecanismos de publicação eletrônica de teses e dissertações, uma interface de busca e o protocolo de OAI-PMH, com vistas à exposição dos metadados para os procedimentos de *harvest*. Esse *software* foi escrito em PHP e utiliza o SGBD MySQL, funcionando tanto em ambiente Linux quanto Windows.

TEDE Simplificado

Considerando as dificuldades de implantação da BDTD Local, apontadas anteriormente, o IBICT promoveu o desenvolvimento do TEDE simplificando os procedimentos de publicação das teses e dissertações. Assim, nessa nova versão, os procedimentos de publicação ficam a cargo da biblioteca, que se responsabiliza por obter a autorização do aluno para publicar a sua tese ou dissertação, e todo o processo de cadastramento e publicação da tese é realizado pela biblioteca. Os resultados obtidos com o repasse dessa nova versão estão se mostrando mais animadores, com melhor receptividade por parte das IES. Essa nova versão utiliza as mesmas tecnologias e plataforma da primeira versão do TEDE.

BDTD / Harvester

Para integrar as várias BDTD locais, o IBICT implementou a BDTD utilizando o SGBD Oracle 9i e a interface de busca em Java. Além disso, o IBICT customizou o *harvester*, que é o mecanismo de *software* responsável por fazer a coleta de metadados junto às IES que implantaram o TEDE ou que habilitaram as suas bibliotecas digitais locais de teses e dissertações com o protocolo OAI-PMH, casos do Inpe e da Unicamp.

O mecanismo utilizado pelo IBICT para implantação da BDTD é a realização de *workshops*. Normalmente o Instituto faz a chamada para esses *workshops* com uma média de 30 a 40 vagas, e as IES interessadas inscrevem, sempre em duplas, um técnico de informática e um de informação, para que a IES tenha total condição para implantar a BDTD local.

Diálogo Científico

Esse projeto tem como base o pacote de *software* E-prints, desenvolvido pela Southampton University. Esse *software* foi customizado para a língua portuguesa pelo IBICT e recebeu a denominação de Diálogo Científico. Trata-se de um *software* que implementa os padrões e ideais da *Open Archives Initiative*. Portanto, é uma ferramenta que permite criar um repositório de dados, portanto, um provedor de dados ao qual o pesquisador pode submeter os seus trabalhos, tais como *pré-prints*, relatórios técnicos, artigos. Além da facilidade de auto-arquivamento, esse pacote de *software* oferece as seguintes facilidades: auto-arquivamento de comentários a um determinado trabalho publicado no repositório, publicação de novas versões de um de-

terminado trabalho, consulta em linha e exposição de metadados para os procedimentos de *harvest* por parte dos provedores de serviços.

Assim, o Diálogo Científico propõe uma nova modalidade de comunicação científica que permite maior interação entre os autores e os leitores, criando, de uma certa forma, um colégio invisível virtual cujos pares estão espalhados pelo mundo integrados por meio da rede Internet.

O IBICT vem estimulando a criação de repositórios de dados nas várias áreas do conhecimento. No caso de as instituições não terem uma infra-estrutura apropriada, o Instituto oferece a possibilidade de hospedar os repositórios utilizando a sua infra-estrutura tecnológica.

Catálogo de Anais de Congresso Eletrônico

Esse subprojeto está sendo realizado pelo Centro de Informações Nucleares da Comissão Nacional de Energia Nuclear (CIN/CNEN). Nesse contexto, o CIN/CNEN customizou o pacote de *software Open Conference System*, o qual foi desenvolvido pela University of British Columbia, no âmbito do projeto *Public Knowledge Project*, coordenado pelo professor John Willinsky. Trata-se de um *software open source* construído em consonância com os padrões e ideais da OAI. Utiliza, portanto, o conjunto de metadados Dublin Core, assim como o protocolo OAI-PMH. Essa ferramenta permite a gestão de todo o processo de organização de um evento científico, oferecendo as seguintes facilidades: criação de página *web* do evento, auto-arquivamento (submissão) de trabalhos, controle do fluxo de avaliação dos trabalhos submetidos (*workflow*), avaliação desses trabalhos e publicação dos anais eletrônico do evento na *web*.

Na adoção dessa ferramenta pelos organizadores de eventos científicos, estes desempenharão o papel de provedores de dados e o CIN/CNEN desempenhará o papel de provedor de serviço, por meio da criação e manutenção automática de um Catálogo de Anais de Congresso Eletrônico, conforme apresentado no esquema da figura 1.

O CIN/CNEN atribuiu a essa ferramenta a denominação de Sistema On-Line de Acompanhamento de Conferências (SOAC).

Integração de Recursos de Informação

A integração de recursos de informação será realizada por meio de uma interface que está em desenvolvimento, utilizando o pacote de

software METALIB. Essa ferramenta de *software* possui os protocolos: i) Z39.50, o qual permite a busca distribuída em várias fontes de informação, em especial as OPACs; ii) OAI-PMH, o qual permite realizar o a coleta (*harvest*) de metadados, consolidando-os em uma base de dados que fica disponível para consulta em linha; iii) http, protocolo utilizado para navegação na *Web*. Essa última modalidade de protocolo exige programação de *scripts* para que se possa simular um usuário entrando em uma página *Web* para fazer uma determinada consulta em uma determinada fonte de informação. Essa programação é feita em linguagem Perl.

A idéia, portanto, é criar uma interface para facilitar ao usuário o acesso às fontes de informação, sem que ele tenha de entrar em cada uma delas pessoalmente, ou seja, é o próprio Metalib que fará a submissão das consultas de um determinado usuário às fontes de informação por ele escolhidas.

8. Considerações finais

Além dos objetivos estabelecidos para a BDB, esse projeto busca a inovação tecnológica e a consolidação da competência nacional no desenvolvimento e uso das novas tecnologias da informação e da comunicação. A BDB conta com o apoio da Finep, que concedeu um financiamento para a aquisição de uma infra-estrutura tecnológica (*hardware* e *software*) capaz de suportar as aplicações planejadas, assim como para o desenvolvimento das tecnologias necessárias e seu repasse às instituições participantes.

Esse projeto propiciou ao Brasil tornar-se o quinto país em número de repositórios de *Open Archives*, à frente do Brasil estão EUA, Reino Unido, Alemanha e Canadá. No mesmo nível do Brasil está a França, com 18 repositórios. Essa posição poderá ser melhorada, se todas universidades que adotaram o TEDE, assim como o protocolo OAI-PMH, inscreverem-se no *ranking* (http://archives.eprints.org/eprints.php?action=browse). Além da iniciativa de teses e dissertações, diversas revistas brasileiras vêm adotando o SEER. Nesse caso, o Brasil se colocará na terceira ou segunda posição.

Além das ações inicialmente planejadas, o IBICT vem trabalhando em outras ações que foram concebidas mais recentemente, motivadas pelas atividades de prospecção tecnológica, as quais possibilitaram a identifi-

cação de outros pacotes de *software*, como o DSPACE, o FEDORA e o CDSware. A identificação e absorção dessas tecnologias permitirão ao IBICT cumprir a sua principal missão, que é a de registrar e disseminar a produção científica e tecnológica brasileira.

A utilização, por exemplo, do DSpace ou do Fedora pelas bibliotecas universitárias com o propósito de abrigar a produção técnico-científica da sua respectiva universidade possibilitará, no futuro, a geração de um catálogo contendo toda a produção técnico-científica brasileira. Por outro lado, o uso intensivo do SEER, pacote de *software* customizado a partir do *Open Journal Systems*, certamente dará origem a um catálogo nacional de publicações seriadas. Essa iniciativa, combinada com articulações políticas que recomendem a publicação preferencial, por parte dos pesquisadores, em repositórios *Open Archives*, poderá tornar o Brasil mais independente das publicações e repositórios estrangeiros. Essa é uma grande oportunidade para países, como o Brasil, despontarem como um dos países produtores de informação, e não apenas consumidor como o era no século passado. Essa autonomia representará uma grande economia para o país.

Dessa maneira, a identificação, absorção e capacitação técnica no uso e desenvolvimento de sistemas de informação com base nos padrões e ideais da OAI constituiu o grande aprendizado proporcionado pelo projeto Biblioteca Digital Brasileira. Esse projeto, além de cumprir as metas propostas, promoveu a internalização de tecnologias e sua distribuição à comunidade provedora de informação em ciência e tecnologia, além de criar competência técnica no uso e desenvolvimento de tecnologias da informação e da comunicação aplicadas ao tratamento e disseminação da informação.

Notas

[1] Segundo Sompel e Lagoze, participaram da Convenção de Santa Fé dois tipos de atores: *data providers* e *service providers*, os quais são definidos como "a data provider is the manager of an e-print archive, acting on behalf of the authors submitting documents to the archive. As pointed out above, the data provider of an open archive will, at least, provide a submission mechanism, a long-term storage system and a mechanism that enables third parties to collect data from the archive; A service provider is a third party, creating end-user services based on data stored in e-print archives. For instance, a service provider could implement a search engine for mathematical e-prints stored in archives worldwide".

[2] Auto-submissão é uma facilidade oferecida pelos sistemas que permite aos autores submeterem os seus trabalhos diretamente nos provedores de dados.

[3] Entende-se por *software open source* aquele cuja distribuição acompanha o seu código fonte. Isso significa que o usuário tem, além do código de máquina do *software*, o seu código fonte, o que lhe dá condições para alterar e adequar o *software* segundo as suas necessidades e eventualmente distribuí-lo. Normalmente esses pacotes de *software* são livres de custo, mas eventualmente eles podem ter um custo associado, segundo os serviços que são oferecidos vinculados ao *software*.

Referências

BOLLACKER, Kurt D.; LAWRENCE, Steve; GILES, C. Lee. CiteSeer: An Autonomous Web Agent for Automatic Retrieval and Identification of Interesting Publications. *SmealSearch*. Disponível em: http://smealsearch2.psu.edu/cache/papers/Business/594 http:zSzzSzwww. csd.uch.grzSz~markatoszSzpaperszSzACM98. Autonomous.Agents.CiteSeer.pdf/DID.pdf/. Acesso em: 21 mar. 2005.

BRIN, Sergey; PAGE, Lawrence. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *CiteSeer*. Disponível em: http://citeseer.ist.psu.edu/brin98anatomy.html. Acesso em: 21 mar. 2005.

CHRISTIAN, Eliot J. GILS: What is it? Where's it going? *D-Lib Magazine*, Virginia (USA), december, 1996. Disponível em: http://www.dlib.org/dlib/december96/12christian.html. Acesso em: 21 mar. 2005.

O'NEILL, Edward T.; LAVOIE, Brian F.; BENNETT, Rick. Trends in the Evolution of the Public Web. *D-Lib Magazine*, Virginia (USA), v. 9, n. 4, april, 2003. Disponível em: http://www.dlib.org/dlib/april03/lavoie/04lavoie.html. Acesso em: 12 mar. 2005.

STAPLES, Thorton; WAYLAND, Ross; PAYETTE, Sandra. The Fedora Project:

An Open-source Digital Object Repository Management System. *D-Lib Magazine*, Virginia (USA), v. 9, n. 4, p. -, April, 2003. Disponível em: http://www.dlib.org/dlib/april03/staples/04staples.html. Acesso em: 12 mar. 2005.

SMITH, MacKenzie e outros. DSpace: An Open Source Dynamic Digital Repository. *D-Lib Magazine*, Virginia (USA), v. 9 n. 1, p. -, January, 2003. Disponível em: http://www.dlib.org/dlib/january03/smith/01smith.html. Acesso em: 12 mar. 2005.

TRISKA, Ricardo; CAFÉ, Lígia. Arquivos abertos: subprojeto da Biblioteca Digital Brasileira. *Ciência da Informação*. vol.30, nº.3, Brasília, Set./Dec. 2001. Disponível em: http://www.ibict.br/cienciadainformacao/viewarticle.php?id=241&layout=abstract. Acesso em: 21 mar. 2005.

WILLINSKY, John; WOLFSON, Larry. The indexing of scholarly journals: A tipping point for publishing reform? *Journal of Electronic Publishing*, v. 7, n.2, 2001. Disponível em: http://www.press.umich.edu/jep/07-02/willinsky.html. Acesso em: 22 mar. 2005

WILLINSKY, John. Proposing a Knowledge Exchange Model for Scholarly Publishing. *Current Issues in Education*, v. 3, n. 6, 2000. Disponível em: http://cie.ed.asu.edu/volume3/number6/. Acesso em: 23 mar. 2005.

VAN DE SOMPEL, Herbert; LAGOZE, Carl. The Santa Fé Convention of the Open Archives Iniciative. *D-Lib Magazine*, Virgínia (USA), v. 6, n. 2, p-, 2000. Disponível em: http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.htm, Acesso em 11 mar. 2005.

Uma biblioteca digital em multimídia para apoiar a educação no Brasil

Fredric Michael Litto *

A educação a distância no Brasil cresce aos trancos e barrancos em quase todos os setores educacionais e está usando todos os sistemas tecnológicos de apresentação – material impresso, rádio, televisão, áudio e videocassetes, cd-roms, Internet e videoconferência. A única exceção se refere à educação superior, uma vez que uma política centralizada de certificação de âmbito nacional permite que forças conservadoras da comunidade educacional que relutam em reconhecer as realizações recentes de educação a distância reprimam seu crescimento.[1] Com o tempo, essa situação mudará, com certeza, para melhor. Mas, até lá, instituições de ensino superior no país continuarão a ganhar experiência com os problemas envolvidos na administração de programas de ensino a distância, embora o façam em a área de legalidade um tanto sombria, uma vez que práticas realmente comuns em outros países são proibidas no Brasil por regulamentos federais. A complexidade no controle de atividades educacionais num país tão grande em área (8.511.965 km^2) em população (180 milhões de habitantes) torna a "calibração" do que realmente acontece no território nacional extremamente difícil, se não impossível.

O Brasil tem um número reduzido dos seus jovens matriculado no ensino superior: aproximadamente 3.5 milhões de estudantes (cerca de 100

* Professor titular aposentado da Escola de Comunicações e Artes da Universidade de São Paulo, coordenador científico da "Escola do Futuro da USP" e presidente, desde 1995, da ABED-Associação Brasileira de Educação à Distância(ABED).

mil dos quais no nível pós-bacharelato). Uma vez que não existem recursos financeiros para a construção de uma centena de *campi* necessários para a expansão da educação superior no Brasil, o país precisaria investir em uma "universidade aberta a distância". Mas 30 tentativas de criar, nos últimos 30 anos, uma tal instituição, de tamanho sucesso em outras partes do mundo, falharam, quando ainda no papel. Há, sem dúvida, uma necessidade para investimento na informação e na tecnologia da informação em todos os níveis de educação no Brasil para superar esses problemas. Nas 1.900 instituições de ensino de terceiro grau, as bibliotecas têm um total de 32.2 milhões de volumes (não computando duplicatas de uma mesma obra, geralmente adquiridas para compensar a impossibilidade de aquisição pelos estudantes), e, quando dividimos esse número pelo total de estudantes, chegamos a uma chocante baixa média de 9.2 livros por estudante. Coincidentemente, havia, em 2004, um total de 367.813 computadores conectados à Internet nas mesmas instituições, dando uma média de 9.5 estudantes por computador. Se não houvesse computadores conectados à Internet nas casas de alguns estudantes e instrutores, seria apropriado perguntar se o Brasil espera entrar de fato na sociedade do conhecimento nesse século.[2]

Estudos recentes da Unesco e associações de editores locais confirmam a situação calamitosa de outros tipos de bibliotecas e livrarias no Brasil. Chamado de "patinho feio" do sistema escolar, a falta de bibliotecas escolares, suplantada somente pela falta de laboratórios de ciências, é predominante em todo o país. Bibliotecas públicas são também uma grande incógnita, com estatísticas governamentais afirmando a existência de cerca de 4 mil dessas instituições, mas não oferecendo informações sobre os critérios de inclusão, se 300, 3.000, 30.000 ou mais livros. A maioria das autoridades sugere, informalmente, que talvez não haja mais que 250 instituições com acesso público livre a coleções com mais de 30 mil volumes. Oitenta porcento dos usuários de bibliotecas públicas, de acordo com as estatísticas, são estudantes, e a maioria das instituições não possui fonte regular de recursos para aquisição, dependendo, principalmente, de doações, e estão acostumadas a um estado constante de desatualização. Finalmente, há em todo o país somente 1.500 livrarias (o ideal seria 10 mil, segundo estudos da Unesco), concentradas em grandes cidades. De fato, 89% dos municípios brasileiros não têm nenhuma livraria.[3]

Há cerca de 22 milhões de computadores pessoais em uso no Brasil em 2004, o que significa que cerca de 12% da população os usa, e há aproximadamente 14 milhões de usuários da Internet (8% da população), somente 816 mil dos quais tem acesso à banda larga, e há 2.2 milhões de *hosts* na *Web*. Esses números colocam o Brasil bem atrás da maioria das nações desenvolvidas e resultam de dois fenômenos: a baixa renda da maioria da população brasileira e o fato de o português, sua língua nacional, representar somente 0,7% do conteúdo da *Web*, oferecendo, assim, pouco incentivo para acesso.[4]

Essa "letargia" se reflete também na presença, ou falta de presença, de computadores e da Internet em escolas de educação primária e secundária. Há um total de 176.880 escolas públicas e privadas no país, com um total de 48.5 milhões de estudantes (o dia escolar é tipicamente apenas de quatro horas de duração, e muitas escolas públicas funcionam em três ou quatro turnos). 80% (141.508) dessas escolas têm eletricidade, e 50% (88.887) têm uma linha telefônica. Vinte e sete porcento (48.671) das escolas têm computador (um total de 276.988 máquinas, 80% deles em escolas particulares) e 11% (19.434) têm acesso à *Web*). Isso representa uma média de 5.7 computadores em cada escola, e 174 estudantes por computador.[5] Examinando esse triste cenário, fica claro que a maioria dos investimentos em tecnologia e educação tanto no setor público quanto no privado está concentradas em *hardware*, *software* e treinamento de professores, e não em conteúdo, o que seria a força motriz para justificar a aquisição de computadores e da Internet em educação em todos os níveis. Sem bom material (bom no sentido de útil para o aprendizado), apresentado com bom uso da língua local, e não somente *chats* e outros desvios dos modelos padrão de escrita) não podemos esperar que exista motivo para a maioria das pessoas se preocupar com computadores e a Internet no processo de aprendizagem.

Por isso, foi com grande satisfação que um grupo de investigadores da Escola do Futuro, laboratório de pesquisa interdisciplinar da Universidade de São Paulo, recebeu, em abril de 1997, a notícia de que seu pedido de subvenção da Fundação AT&T em Nova York para a criação de uma "biblioteca virtual para o estudante brasileiro" tinha sido aprovado. A subvenção, decidida por um seleto painel de especialistas organizado pelo International Council for Open and Distance Learning (ICDE) forneceu US$100 mil para o projeto e pouco depois, outra subvenção de US$

40 mil da Secretaria de Cultura do Estado de São Paulo permitiu o início de um processo de "inclusão cultural" que continua até hoje, início de 2005. Os passos iniciais foram uma demonstração clara da observação hoje corrente de que, se no passado gastavam-se 20% do tempo de um projeto no seu planejamento e 80% em seu desenvolvimento, hoje em dia, devido à grande complexidade que nos cerca, o reverso é, na maioria da vezes, o caso: 80% para planejamento e meramente 20% para execução do que foi planejado. O grupo de pesquisa levou mais de um ano para organizar os aspectos documentais, tecnológicos e dos fluxos das tarefas antes de efetivamente "entrar no ar".

As tarefas iniciais foram as seguintes: 1. formação do time de pesquisadores; 2. fixação de parcerias internas e externas; 3. estabelecimento de princípios para negociar questões sobre direitos de propriedade intelectual; 4. aquisição de *hardware*, *software* e outros materiais; 5. criação de espaços de trabalho e rotinas que conduzem à produtividade; 6. construção de interface do usuário; 7. iniciação de busca de patrocínio para o projeto após april de 1998. A equipe era formada por um coordenador geral, um especialista em interfaces na rede, um especialista em *design* gráfico, um coordenador de produção e vários assistentes de iniciação científica para a digitação de dados, correspondência com usuários e outras tarefas importantes. Teve início também a troca de idéias com a instituição associada, parte da proposta original, a University of British Columbia, do Canadá, na pessoa do doutor Tony Bates, renomado estudioso e profissional de educação a distância, que concordou, gentilmente, em servir como avaliador formativo do projeto.

Os objetivos gerais iniciais do projeto: 1. fornecer textos completos de obras clássicas e de difícil aquisição da cultura brasileira adequados para estudantes do ensino básico (obras literárias e históricas, obras de referência, atlas, imagens, sons e vídeos); 2. aperfeiçoar as habilidades heurísticas de estudantes e professores; 3. aumentar o desejo de implantar computadores e acesso à rede nas escolas; 4. estudar o verdadeiro uso no dia-a-dia da biblioteca virtual pela sua comunidade de usuários; 5. investigar os aspectos de custo-benefício de uma tal coleção de materiais em multimeios que oferece apoio ao ensino a distância formal ou informal.

Uma importante parceria "interna" foi firmada com o Museu de Arqueologia e Etnologia da Universidade de São Paulo, que cedeu para a

Biblioteca Virtual imagens valiosas de cerâmica, cocares e outros artefatos culturais dos povos indígenas brasileiros. Outra parceria externa importante foi firmada com a Fundação Roberto Marinho do Rio de Janeiro, em sociedade com a Federação das Indústrias do Estado de São Paulo (Fiesp) e a agência de nível nacional Serviço Social da Indústria (Sesi), que, juntos desenvolveram o Telecurso 2000, cursos completos do ensino fundamental e médio, distribuidos por meio de programas de televisão em circuito aberto e materiais impressos disponíveis em bancas de jornais em todo o país.[6]

Criado principalmente para adultos já envolvidos em atividade econômica, mas com escolaridade incompleta (o trabalhador médio no Brasil ainda tem apenas seis ou sete anos de educação formal) e desejosos de completála por meio de ensino a distância em tempo parcial, o programa Telecurso 2000, em 2004, teve mais de 600 mil alunos matriculados. Com ajuda do superintendente do programa, professor Marcos Formiga, a Fundação Roberto Marinho deu à Biblioteca Virtual do Estudante Brasileiro permissão para usar todo o material do curso já em forma digital, que continua sendo, até hoje, uma das partes mais acessadas da Biblioteca.

Outra colaboração foi firmada com a Vozoteca (biblioteca de vozes) de Luiz Ernesto M. Kawall, um colecionador independente de sons do passado do Brasil, especialmente gravações de personalidades políticas da primeira e segunda metades do século XX. Essas gravações disponibilizadas digitalmente pela Biblioteca Virtual permitem aos estudantes ir além do livro de texto tradicional, ouvindo, de fato, os estilos oratórios de figuras históricas cujos nomes estão constantemente na sua presença em livros de texto, nomes de ruas e mesmo nas palavras de canções populares.

Outros planos foram feitos, mas, devido às dificuldades de financiamento, nunca foram implantados. Talvez valha a pena listá-los aqui como sugestão para outros que porventura queiram criar bibliotecas virtuais desse tipo. Esperávamos originalmente, por exemplo, ter a informação básica na Biblioteca Virtual em português e navegação dentro da base de dados em português, inglês, espanhol e, talvez, até mesmo em francês, permitindo assim, aos estudantes de diferentes partes do mundo que se interessam por Brasil e sua cultura, acesso ao material. Mas os recursos disponíveis para fazer isso nunca foram suficientes. Igualmente, espera-

va-se criar uma *interface* para usuários com três eixos: localização (mapa mundi – ao clicar em algum lugar, aparece tudo o que a ele se refere na base de dados); tempo (uma barra de tempo deslisante); tópico (palavras-chave de acesso a todos os arquivos), mas isso não foi possível. A equipe também não foi capaz de dar início a um setor Frequently-Asked Questions (FAQ), capaz de permitir aos usuários obter muitas respostas a suas questões referentes à Biblioteca.

Outros problemas de natureza genérica capazes de desencorajar alguns futuros criadores de bibliotecas virtuais foram encontrados: 1. a ausência notável de uma tradição de trabalho de grupo em pesquisa nas humanidades, pelo menos como a vemos no Brasil, constantemente atormentando as tomadas de decisão e a produção; 2. a escolha de técnica de entrada de dados também impediu o desenvolvimento do trabalho – leitura óptica de caracteres de textos se mostrou não ser prática, e a digitalização dobrada de quase todos os textos foi tomada como norma; 3. o problema das muitas mudanças ortográficas que ocorreram na língua portuguesa no correr dos últimos 100 anos e a inabilidade de os jovens estudantes de hoje entenderem textos nos formato ortográfico e gramatical originais exigieram muita reformatação; 4. a existência de menos material em domínio público disponível do que se esperava, obrigando o grupo a buscar novas soluções. Editoras resistiam ao uso gratuito de seu material na Internet, mesmo com o apoio da associações de classe. Em vez da "aceitação por atacado" que nós esperávamos da parte das editoras, com a permissão de colocar na Internet suas obras esgotadas, a equipe teve de se engajar numa "operação varejista", conseguindo, só ocasionalmente, permissão na base de título a título. Esperava-se que um seminário sobre propriedade intelectual, novas tecnologias de comunicação, e a educação, em associação com uma organização nacional de editoras, ia ser benéfico, mas isso ainda não aconteceu, talvez porque os setores implicados ainda não reconhecerem a importância do tema.

Finalmente, esperava-se ter uma rica área de atividades paradidáticas dentro da Biblioteca Virtual, com referência ao material audiovisual e textual dentro da coleção. Tais atividades paradidáticas incluiriam *role-playing games*, troca de informação entre pares, consulta a especialistas importantes pela Internet e consulta a outras bases de dados. Há, na verdade, uma troca de idéias ativa e proveitosa entre professores sobre currículo,

atividades na sala de aula e bibliografia, mas isso ainda não atingiu o estágio de um sistema poderoso de apoio originalmente imaginado.

A Biblioteca Virtual do Estudante Brasileiro está disponível gratuitamente a todos os interessados 24 horas por dia, sete dias por semana, por meio do World Wide Web (www.bibvirt.futuro.usp.br). Seu conteúdo está dividido em três seções: a coleção, atividades e sugestões para fontes de pesquisa. A coleção consiste de textos completos de obras literárias e documentos históricos, artigos de periódicos científicos, imagens, sons e *software*. Mais de 300 obras da literatura brasileira em domínio pública estão disponíveis pelo índice de autor, título e gênero, e muitas delas são "baixáveis" para computadores e *palmtops*. Resenhas de muitas dessas obras estão também disponíveis, bem como cinco obras completas da literatura internacional traduzidas para o português (Lewis Carrol, Alexandre Dumas Fils, Georg Büchner, and Henry Rider Haggard). Há, também, um *link* para a célebre biblioteca virtual Gutenberg, uma coleção digital existente há 30 anos, e possuindo agora mais de 10 mil obras literárias, em 16 diferentes línguas, todas elas em texto integral. A seção de imagens (a maioria em formato JPEG) inclui muitos dos desenhos abundantes em cores de Jean Baptiste Debret, o francês que visitou o Brasil no período de 1816-1831 e deixou um rico relato pictórico da vida econômica e social na sociedade colonial. Desenhos dos diferentes pássaros do Brasil e fotografias da enorme variedade de frutas do Brasil estão presentes, bem como uma grande coleção de *clipart* especialmente organizada para jovens estudantes. A seção de sons (em Real Audio, MP3, MIDI e RAM) deve muito à colaboração de Luiz Ernesto M. Kawal, que forneceu as gravações de eventos históricos, personalidades e políticos do passado, bem como famosos *jingles* e *spots* de publicidade da mídia de décadas atrás. Usuários podem escolher entre gravações que registraram a primeira voz gravada no Brasil, vozes das revoluções de 30 e 32, da II Guerra Mundial, da História do Rádio no Brasil, Brasil nos campeonatos mundiais de futebol, compositores de Bossa Nova, e a história do samba e de famosas escolas de samba. Uma área divertida de pesquisa musical oferece fotografias, partituras e a gravação de sons de 14 diferentes instrumentos de percussão do Brasil. Há, também, uma pequena (mas em crescente desenvolvimento) coleção de "livros falados", incluindo muitas obras de literatura, preparada para pessoas com deficiência visual, e disponível para todos os interessados,

graças à cooperação da Fundação Dorina Nowill e da Auditeca Kaete Heymann da Comunidade Shalom de São Paulo.

Uma considerável coleção de revistas científicas, difíceis de serem encontradas, está também disponível, cedida pelos editores interessados em ajudar futuras gerações de estudantes: *Revista Brasileira do Ensino de Física*, *Revista do Professor de Matemática*, *Física na Escola*, *Revista do Departamento de Geografia*, *Educação e Pesquisa*, *Sinopse – A Revista do Cinema* e outras.

Entre "materiais especiais", encontram-se tesouros como uma coleção de mais de 80 entrevistas gravadas no período de 1984-1989 com importantes cientistas brasileiros, e várias dezenas de entrevistas "baixáveis" de vídeo com importantes figuras como Paulo Freire (suas entrevistas são um absoluto *hit*, recebendo um milhar de "baixadas" por semana). Alberto Dines, Juscelino Kubtschek e líderes das tribos Araweté e Krenakarore da região amazônica, entre outros. Espalhado por toda a coleção, encontra-se um amplo material representando ricas tradições folclóricas brasileiras em textos, imagens e sons.

A área reservada para "atividades" enfatiza interatividade com os usuários, bem como entre os próprios usuários, especialmente com experiências relacionadas ao uso da Biblioteca Virtual no processo ensino/aprendizagem. Há um "quadro de mensagens" para anúncios e trocas de idéias, sugestões, noticiário de eventos, um calendário cultural, e uma Caça de Tesouro Online Anual, que estimula estratégias de busca criativa de informação dentro da Biblioteca Virtual e resulta em premiações (geralmente *palmtops* e livros doados por livrarias locais). Etcetera é uma seção que cita fontes adicionais de pesquisa disponíveis na Web, bem como os endereços de bibliotecas e livrarias reais no país, complementando as fontes disponíveis na Biblioteca Virtual.

Vistos como um todo, os conteúdos da Biblioteca Virtual refletem seu desenvolvimento: com exceção do extensivo e bem organizado material do Telecurso 2000, o restante é, mais precisamente, uma maravilhosa miscelânea de conhecimento, informação e sabedoria incomum. Devido à falta de financiamento consistente desde o término das doações iniciais, fomos tolhidos por severas restrições sobre direitos autorais. Mesmo assim, acabamos gostando do crescimento não-linear, exponencial e eclético em várias direções. Mas, responsáveis pelo desenvolvimento da Bibliote-

ca Virtual, assumimos seus defeitos e falhas e deleitamo-nos com o grande uso dado, apesar da irregularidade de seu acervo. Esperamos pelo dia em que as virtudes da Biblioteca Virtual e sua importância estratégica sejam reconhecidas por agências financiadoras generosas, que permitirão realizar os melhoramentos necessários. Impacientes, contudo, como somente os membros da equipe de uma nova biblioteca podem ser, envolvemo-nos em ação para mudar o nome Biblioteca Virtual do Estudante Brasileiro para Biblioteca Virtual do Estudante da Língua Portuguesa, pois assim será possível crescer mais, incluindo material educativo dos oito países na Europa, América do Sul, África e Ásia, onde o português é a língua principal, recebendo de, e reencaminhando para cada país, material para o ensino e o entretenimento que poderá ser entendido por todos.

Em 2001, Luciana Salgado, uma estudante no Programa de Mestrado da Escola de Comunicações e Artes da Universidade de São Paulo, fez um estudo de alguns aspectos da Biblioteca Virtual. Sua tese tentou determinar se a Biblioteca Virtual havia atingido seus objetivos originais e qual o perfil de seus usuários.[7] Em 2001, o número médio diário de usuários discretos excedeu 5 mil. No decorrer dos três anos subseqüentes, a Biblioteca Virtual recebeu seis troféus "I-Best" votados pelos internautas brasileiros por excelência no conteúdo nas categorias de educação e treinamento, e arte e cultura, e seu número diário de usuários cresceu para 15 mil. Contudo, os resultados da investigação de 2001 merecem atenção.

De setembro de 2000 a 25 de fevereiro de 2001, um questionário contendo 37 itens preparado por Salgado tentando determinar o perfil de usuários e a natureza de sua satisfação ou insatisfação foi colocado no *site* da Biblioteca Virtual, obtendo um total de 528 respostas, das quais 479 foram efetivamente usadas para o estudo. Os resultados do levantamento trouxeram algumas surpresas: para um *site* planejado para estudantes de primeiro e segundo graus, os usuários eram um tanto mais velhos: 10-13 anos de idade, 7.0%; 14-17 anos, 20.9%; 18-21 anos, 17.4%; 22-15 anos, 15.1%; 26-29 anos, 20.9%; 30-39 anos, 18.3%; acima de 40 anos 11.7%. Usuários morando em capitais de estados eram 50.8%, e os do interior, 49.2%. Usuários residentes em cidades com população superior a 1 milhão de habitantes eram 43,0% do total, enquanto os moradores de cidades acima de 500 mil habitantes eram 14%, habitantes de cidades entre 500 mil e 100 mil habitantes, 22.0%, os de cidade com

menos de 100 mil habitantes, 14.0%, e os de cidades com menos de 10 mil habitantes, 7.0%. Estudantes de escolas primária e secundária abrangiam 37,8% dos usuários, enquanto estudantes de nível universitário abrangiam 23,8% do total. Estudantes do sexo masculino eram 50.7% e do feminino 49.3%. Estudantes de escolas públicas representavam 50,7% do total e os de escolas privadas 49,3%. Estudantes do Estado de São Paulo constituíam 43,0% dos usuários, os de Minas Gerais 10,0%, os do Rio de Janeiro, 8,0%, e os do Paraná, 7,0%. A renda mensal familiar informada foi bem mais alta do que o esperado: até R$300,00, 3,3%; R$ 301,00 a R$ 600,00, 3,3%; R$ 601,00 a R$ 1.200,00, 14,7%; R$ 1.200,00 a R$ 3.000,00, 17,45; R$ 3.001,00 a R$ 6.000,00, 3,3%; R$ 6.001,00 a R$ 10.000,00, 11,2%; R$ 10.001,00 a R$ 20.000,00, 17,4%; acima de R$ 20.000,00, 10,4%.[8] Ao serem perguntados de onde acessavam a Internet, os usuários responderam o seguinte: de suas casas, 72,1%; do trabalho, 17,4%; da escola 5.4%; da casa de amigos ou parentes, 3,1%; da biblioteca 1,7%. Também, 84,4% responderam que usaram a Biblioteca Virtual para estudo e pesquisa, enquanto 8.4% indicaram prazer, e 7.2%, relação com o trabalho.

Talvez digno de preocupação e atenção no futuro seja o fato de 57,3% terem respondido que seus professores pediram pesquisa na *Web*, mas não deram orientação; 25,4% disseram que os professores deram uma orientação mínima; 17,3% indicaram que seus professores deram orientação inicial para sua pesquisa na *Web*. Quando perguntados se seus professores os encorajavam a fazer pesquisa usando a Internet, houve respostas diferentes da parte de estudantes de escolas públicas e privadas:

"Todos os seus professores encorajaram sua pesquisa na *Web*?": Dez por cento dos estudantes das escolas públicas deram resposta positiva, enquanto somente 8,0% dos estudantes de escolas privadas responderam afirmativamente; "Alguns professores?" 50,0% dos estudantes de escolas públicas e 44,0% dos de escolas privadas confirmaram positivamente; "Nenhum professor?"; 40,0% dos estudantes de escolas públicas e 48,0% dos de escolas privadas responderam positivamente.

Em relação à general satisfação geral dos usuários com a Biblioteca Virtual, 57,0% responderam afirmando que encontraram exatamente o que estavam procurando, enquanto 44,1% afirmaram que "normalmente" encontravam o que procuravam. À pergunta sobre se foram capazes

de localizar o material desejado dentro da biblioteca: 34,4% responderam "sempre"; 44,4% responderam "quase sempre"; 15,7% responderam "algumas vezes"; 8,4% responderam "raramente"; e 7,45 responderam "nunca". Se ou não o material encontrado na Biblioteca correspondeu às necessidades de sua pesquisa: 57,0% responderam que encontraram "exatamente o que procuravam"; 12,1% acharam o material "muito sofisticado"; e 30,9% acharam o material "um tanto simplório". Se achavam que o material procurado na Biblioteca Virtual poderia ser encontrado facilmente em outras fontes, 36,9% disseram que isso seria possível, 33,8% disseram que "às vezes achavam isso" ser verdadeiro, 13,6% disseram que era difícil, e 3,5% disseram que nunca era o caso.

Na ordem de importância para os usuários, o acervo da Biblioteca foi assim avaliado: literatura, material didático, material paradidático, sons e imagens. Com relação à navegação dentro da Biblioteca, 90,4% relataram não ter tido dificuldade, 9,4% sentiram dificuldade; 48,5% relataram usar o Internet Explorer 5x, enquanto 19,2% responderam usar Internet Explorer 4x. Em relação à forma em que eles leram o material encontrado na biblioteca, 28,0% leram diretamente na tela; 13,4% imprimiram diretamente enquanto se achavam ligados à Internet; 34,3% "baixaram" para leitura posterior *on-screen*, e 23,2% baixaram para posterior impressão e leitura. Salgado concluiu que o público destinatário original da Biblioteca Virtual tinha na verdade sido alcançado, mas o aspecto de "acesso livre e irrestrito à *Web*" tornou o mesmo conteúdo disponível para um público muito maior do que o originalmente visado. Estudantes universitários e vestibulandos parecem ter usado a Biblioteca Virtual como uma maneira de ter acesso "sem-livros" às leituras exigidas para as quais sua fonte financeira era inadequada. Ela também concluiu que, embora melhorias pudessem ser feitas em relação à interface entre o usuário e o sistema, em geral, os usuários pareciam satisfeitos com a estrutura e a natureza do acesso ao conteúdo da Biblioteca.

O número reduzido de bibliotecas escolares, bibliotecas públicas e livrarias no Brasil, claramente incompatível com as necessidades de uma economia baseada em conhecimento, e as verbas públicas disponíveis no passado e aparentemente no futuro para retificar essa situação obrigam aqueles preocupados com a produtividade futura do Brasil e sua capacidade de competir globalmente a experimentar com soluções para

problemas sociais baseadas nas novas tecnologias da comunicação. A maior parte da informação na *Web* está em inglês, portanto com acesso limitado para estudantes de países com outras línguas principais. Países em desenvolvimento precisam usar novas tecnologias do tipo usado no projeto da Biblioteca Virtual do Aluno de Língua Portuguesa, quer vista como um *site* independente na *Web* ("*stand alone*"), quer como "sistema de apoio" para estudantes matriculados em cursos *on-line*. Os benefícios a seram atingidas são os seguintes: (1) dar ao estudante acesso à sua própria literatura nacional, iconografia e sons do passado; (2) estimular a leitura e a pesquisa em coleções de fontes nacionais online; (3) ajudar a justificar o custo da compra e manutenção de computadores e acesso à *Web* em escolas e centros comunitários, oferecendo, em língua local, informação útil para o aprendizado. Isso, por sua vez, diminui as diferenças entre o aprendizado nas zonas rural e urbana e promove, em geral, uma melhor e sustentável qualidade de vida.

Notas

[1] Litto, Fredric (2002). The Hybridization of Distance Learning in Brazil — An Approach Imposed by Culture. *International Review of Research in Open and Distance Learning:* 2, 2 [iuicode: http://www.icaap.org/iuicode?149.2.2.6 Accessed 15 April 2004].

[2] Para estatísticas tratando da população, indicadores sociais nacionais, educação e condições de vida, ver o *site* do Instituto Brasileiro de Geografia e Estatística [www.ibge.gov.br Acessado 15 de maio de 2004.] Para estatísticas mais específicas sobre educação fundamental e média ver *Dataescolabrasil; Edudatabrasil; Censo Escolar 2004*; e *Sinópse Estatística da Educação Básica 2003*, e educação pós-secundária *Sistema de Informação* da Educação Superior (SiedSup); Cadastro de Cursos e Instituições *da Educação Superior*, ver o *site* do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira [www.inep.gov.br Acessado 19 de maio de 2004]. Ver também: Fredric M. Litto, org. *Campus Computing Report.Br 2004. Computação e Tecnologia de Informação nas Instituições de Ensino Superior no Brasil*. São Paulo: Altana, 2005.

[3] Sobre bibliotecas e livrarias ver Cecília Jorge, "Brazil, a Nation that Doesn´t Read" *Brazzil Culture*, June 2004 [www.brazzil.com/2004/html/articles/jun04/p109jun04.htm Acessado 17 de maio de 2004.]; Ottaviano De Fiori, "As políticas do livro" [www.minc.gov.br/textos/olhar/politicaslivro.htm Acessado 15 de maio de 2004]; "Estudo revela que cerca de mil cidades brasileiras não têm bibliotecas" [26/4/04] Notícias de 17 de junho de 2004. *AjudaBrasil* [www.ajudabrasil.org/noticias.asp? idnoticia=50]; e Joseana Paganine, "Câdê o Leitor?" [<www2.univille. edu.br/biblioteca/boletim_junho2002/PAGANINE. htm> Acessado 17 de junho de 2004].

[4] *Informativo INEP*, Ano 2, No 38, 11 Mai 2004 [www.inep.gov.br/informativo/informativo38.htm Acessado 17 de junho de 2004].

[5] Marcos Dantas Loureiro, secretário de Educação a Distância, Ministério da Educação do Brasil, "O Ministério da Educação e a EAD: Visão Geral em, Legislação Atual" Seminário Internacional de Educação a Distância, Brasília, UNILEGIS, 9 de junho de 2004. Ver também Fredric M. Litto (1998), "Culture and Entropy at the Interface of Freedom of Expression and the New Communications Technologies" in *Freedom of Expression and the New Communications Technologies*, eds. Michèle Paré and Peter Desbarats (Montreal, Canadá: UNESCO and IQ Coletif), pp. 201-09.

[6] Oliveira, J.B.A., Moura Castro, C. e Verdisco, A. (2003). Education by Television: Telecurso 2000. In J. Bradley (Ed.). *The Open Classroom. Distance Learning In and Out of Schools*. London and Sterling, VA: Kogan Page, pp.133-145.

[7] Luciana Salgado, "A Biblioteca Virtual do Estudante Brasileiro da Escola do Futuro da Universidade de São Paulo: Um Estudo da sua Estrutura e dos Seus Usuários", Dissertação de Mestrado, Escola de Comunicações e Artes, Universidade de São Paulo, 2001. Ver também Fredric M. Litto, "Digital Libraries, Developing Countries, and Continuing Education", *Continuing Higher Education Review* (Harvard University), Vol. 68, Fall, 2004, pp. 78-86.

[8] Em dezembro, 2000, um dólar norteamericano valia aproximademente R$ 2.00.

Federação de bibliotecas digitais lusófonas em ciências da comunicação

*Sueli Mara S. P. Ferreira**
*Patrícia Cristina N. Souto***

1. Introdução

A busca por maior efetividade na tomada de decisões com base em conhecimento científico tem sido progressivamente influenciada pelo amadurecimento na inserção e adoção das tecnologias da informação e de comunicação. Somado a isso, o acesso, uso e compartilhamento de informação técnico-científica qualificada têm viabilizado, agilizado e potencializado os processos de planejamento, gestão, pesquisa e ensino em todas as áreas de conhecimento.

Vários desafios, neste cenário dinâmico, contribuem para a exaustão e insuficiência dos modelos atuais de gestão da informação: usuários mais experientes e exigentes; mudanças na comunicação científica; multiplicação de formatos dos ativos digitais nas instituições acadêmicas; surgimento da Iniciativa de Acesso Aberto[1]; gestão e segurança

*Coordenadora da Portcom/Intercom e do Núcleo de Pesquisa "Design de Sistemas Virtuais Centrado no Usuário". Professora doutora do Departamento de Biblioteconomia e Documentação da ECA/USP. smferrei@usp.br
**Pesquisadora da Portcom/Intercom e do Núcleo de Pesquisa "Design de Sistemas Virtuais Centrado no Usuário". Aluna de mestrado do Programa de Pós-Graduação em Ciências da Comunicação, área de concentração em ciência da informação e documentação. patriciacn@uol.com.br

operacional das fontes informacionais; preservação dos objetos digitais; necessidade de replicação de serviços para prover acesso global; o fortalecimento da computação ubíqua e dos ambientes de sistemas de informação em rede, dentre outros.

As bibliotecas digitais, diante destes desafios, vêm alterando seu escopo e função, transcendendo a ênfase na organização, distribuição e recuperação da informação, passando a se posicionar como um ambiente integrado de informação com foco no apoio aos processos de aprendizagem, a resolução de problemas e a construção de conhecimento individual e coletivo. Feng e outros (2005) acredita "que o futuro das bibliotecas digitais vai além do armazenamento e do registro e passa a ser um ambiente de criação, aquisição, compartilhamento e propagação do conhecimento".

Neste mesmo sentido, Coleman e outros (2002) considera as bibliotecas digitais como sendo um ambiente integrado pela composição de espaços de informação, de aprendizado e de interação. O "espaço de informação" se compõe de coleções (objetos digitais cada vez mais complexos e heterogêneos), serviços (descoberta da informação, busca, identificação, criação de metadados), ferramentas para o uso da informação (anotação, personalização, alertas etc.) e interfaces. O "espaço de aprendizagem" se compõe de estrutura e combinações dos e entre os objetos digitais do espaço de informação com foco no aprendizado do usuário, incorporando a recuperação de informação contextualizada, interativa e controlada de acordo com os estilos de aprendizado dos usuários. O "espaço de interação" proporciona situações e recursos que libertam o usuário das limitações físicas para a interação, proporcionando os benefícios das comunidades e interação virtual (interatividade síncrona e assíncrona) e a construção conjunta de conhecimento.

A operacionalização deste novo conceito de biblioteca digital, mais focada em se colocar como um espaço de aprendizagem e interação, tem sido tema de discussão e reflexão por vários autores. Dentre eles, Feng (2001, 2002, 2004) delineia um modelo de biblioteca digital que envolve duas dimensões que se complementam:

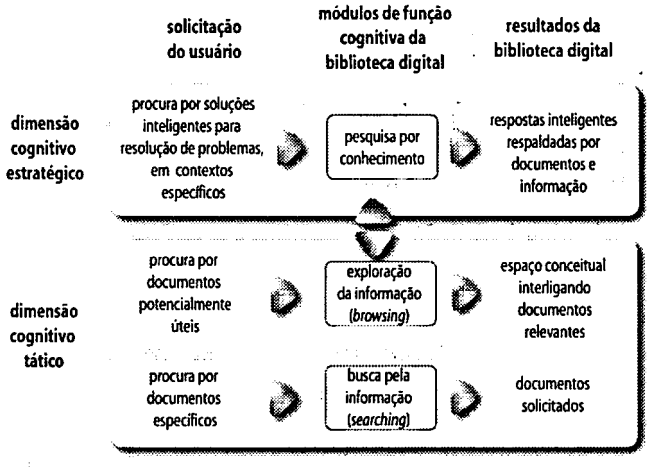


Figura 1 – Modelo de Biblioteca Digital proposto por Feng (2001, 2002, 2004)

- Dimensão "Cognitivo Tático": é entendida como o "espaço documental" onde se priorizam os processos cognitivos táticos (comportamento de busca/*searching* e de exploração/*browsing*) dos usuários proporcionando agilidade e facilidade na recuperação de documentos. Sua base está na gestão documental e informacional e nos processos de armazenamento e distribuição.
- Dimensão "Cognitivo Estratégico": é entendida como o "espaço do conhecimento" onde se priorizam os processos cognitivos estratégicos (busca de soluções inteligentes para problemas informacionais de alta complexidade) dos usuários, proporcionando a correlação e inter-relação de informações em termos semânticos, temáticos, lógicos, físicos, dentre outros.

A viabilização destas novas perspectivas, que expandem a atuação das bibliotecas digitais, só é possível com a convergência e integração das atividades, dos serviços, das práticas, das funções, dos procedimentos, dos conteúdos, das tecnologias, bem como com a participação dos atores envolvidos em cada um dos processos de gestão da informação. Uma das estratégias apontadas pela literatura para atender a estas exigências é federar as bibliotecas digitais.

Segundo Arms (2000), o termo biblioteca digital federada "descreve um grupo de organizações, trabalhando juntas formal ou informalmente, que

concordam em assumir uma série de serviços e padrões comuns, buscando a interoperabilidade entre elas. Os parceiros em uma federação podem ter diferentes plataformas tecnológicas, desde que implementem serviços compatíveis com os padrões determinados. Necessitam concordar tanto com os padrões tecnológicos como com as políticas e diretrizes (incluem normas de direitos autorais, segurança, privacidade e acordos financeiros)".

Pirri e outros (2002) e Gonçalves, France e Fox (2001) complementam este conceito agregando a idéia de que uma Federação se refere à composição ou união de várias bibliotecas digitais independentes ou autônomas, distribuídas na Internet e organizadas por temas ou recursos específicos, que formam uma rede de bibliotecas com interface de acesso único e integrado à grande quantidade e heterogeneidade de informação estruturada, semi-estruturada e não-estruturada (vídeos, apresentações, imagens, sons, gráficos etc.).

A proposição de uma federação de bibliotecas digitais para a comunidade lusófona em ciências da comunicação é resultado do processo evolutivo das atividades, serviços e produtos desenvolvidos pela Portcom / Rede de Informação em Ciências da Comunicação dos Países Lusófonos, criada em 1981 pela Intercom / Sociedade Brasileira de Estudos Interdisciplinares da Comunicação. Este documento apresenta o projeto da federação e detalha o modelo proposto desde sua concepção até sua operacionalização e gestão.

2. Histórico resumido da Portcom

A carência e dispersão de fontes bibliográficas específicas para a área de ciências da comunicação e as dificuldades enfrentadas pelos pesquisadores para ter acesso ao conhecimento acumulado fizeram com que, em 1981, fosse criado pela Intercom um centro de documentação destinado a coletar e processar a produção científica em ciências da comunicação.

Em 1988, apoiado pela Unesco[4], este Centro amplia seu escopo passando a servir de pólo centralizador e gerenciador da produção técnica e científica lusófona. Como decorrência da experiência adquirida e das atividades desenvolvidas, depois de 1999 a Portcom (já postulada como rede de informação) passa a assumir o papel de agente integrador e articulador

de políticas, ações e instrumentos que normalizam e implementam estratégias para o desenvolvimento e uso de recursos e competências lusófonas em ciências da comunicação. Como resposta a este novo papel surge o projeto de criação da federação que ora se apresenta.

3. Projeto da Federação

O projeto da Federação de Bibliotecas Digitais Lusófonas em Ciências da Comunicação engloba as seguintes etapas:

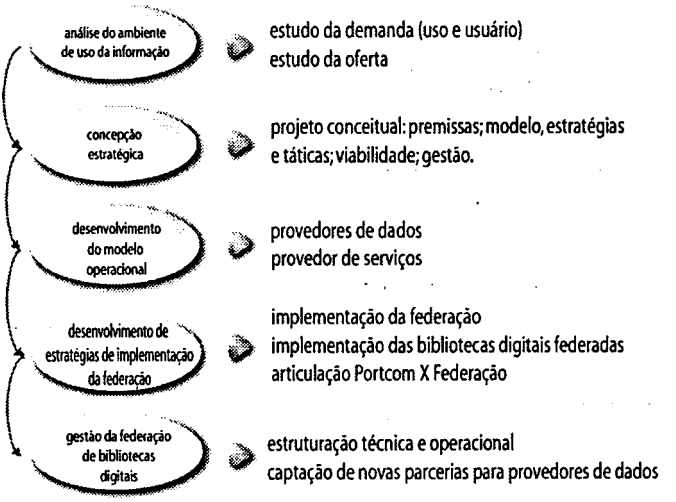


Figura 2 – Etapas do Projeto da Federação de Bibliotecas Digitais Lusófonas em Ciências da Comunicação

Análise do ambiente de uso da informação

Visando diagnosticar, mapear e entender o contexto informacional brasileiro em comunicação, a equipe da Portcom[5] desenvolveu pesquisa junto a 117 bibliotecas e centros de informação de universidades públicas e privadas, cobrindo todas as regiões brasileiras.

Os resultados mais relevantes desta pesquisa apontaram o seguinte: grande concentração de bibliotecas na região Sudeste, seguido da região Sul, Nordeste, Centro-Oeste e Norte; a existência de poucas bibliotecas setoriais dedicadas (ou específicas) a ciências da comunicação; imensa diversidade de serviços e produtos com automação total ou parcial; grande variedade de

softwares em uso; inexistência de tesauro e/ou vocabulário controlado em português em ciências da comunicação; predominância de livros nos acervos; ausência de revistas científicas lusófona indexadas; ausência de publicações não tradicionais (filmes, vídeos etc) no acervo; participação majoritária em redes de serviços e/ou produtos genéricos (tipo Comut e CCN).

Tais resultados, somados a revisão de literatura e análise das ferramentas tecnológicas disponíveis, forneceram as condições necessárias para o detalhamento do projeto conceitual da Federação de Bibliotecas Digitais Lusófonas.

Concepção estratégica da Federação (projeto conceitual)

A Federação é concebida estrategicamente com base no modelo do Feng (2004, 2002, 2001), privilegiando e alinhando sua atuação às duas dimensões propostas – "cognitivo tático" e "cognitivo estratégico". Desta forma, alicerça-se em várias premissas, tais como:

- escopo do público-alvo – atuar com todos os atores envolvidos no processo de gestão da informação técnica, acadêmica e científica, categorizados como produtores, gestores de fontes, gestores de conteúdo, distribuidores de conteúdo e comunidade em geral;
- alinhamento ao público-alvo – se alinhar continuamente às necessidades e comportamentos informacionais dos diferentes públicos-alvo para subsidiar as atividades e processos de tomada de decisão nos âmbitos de planejamento, desenvolvimento, implementação, gestão e evolução;
- acesso público e universal – enfatizar o alinhamento das atividades com a filosofia do Movimento do Acesso Livre e a socialização da informação técnico-científica;
- política de conteúdo – obedecer a critérios, normas e padrões internacionais de qualidade para organização e tratamento do conteúdo. Incorporar conteúdos estruturados e não-estruturados, fontes tradicionais e não-tradicionais, formatos e suportes heterogêneos que favoreçam tanto a dimensão "cognitivo tático" como a "cognitivo estratégico";
- política de serviços: desenvolver serviços que atendam às dimensões "cognitivo tático" e "cognitivo estratégico";
- política de tecnologias – privilegiar a utilização de *software* livre, com base no protocolo OAI-PMH (*Open Archive Initiative-Protocol for Metadata Harvesting*), especialmente customizados para as questões de interoperabilidade:

• gestão da federação – adotar um modelo de gestão compartilhada (descentralização e co-responsabilidade), valorizando as especificidades de cada biblioteca federada, garantindo sua autonomia e criando estratégias de motivação para cooperação contínua. Flexibilizar o desenvolvimento de bibliotecas em torno de eixos: geográfico (estado, região, país), temático (áreas especializadas da comunicação) e processual (metodologia, tecnologia, procedimentos).

Desenvolvimento do modelo operacional para a Federação

O modelo operacional da Federação é determinado pela adoção do protocolo de transferência OAI-PMH, que por sua vez foi selecionado em função dos valores agregados aos parceiros/bibliotecas federadas e aos usuários finais. Para as bibliotecas parceiras, o uso do protocolo garante interoperabilidade[4], proporciona facilidade e reduz custos na implementação. Para os usuários finais, viabiliza, por meio de uma interface única, o acesso e uso integrado de informações e serviços advindos de múltiplas bibliotecas digitais (LIU, 2002; SHI, 2004).

O protocolo OAI-PMH adota o conceito de "provedores de dados (PD)" e "provedores de serviços (PS)". Segundo IBICT (s.d.), os "provedores de dados administram sistemas que mantêm repositório de dados e suportam o Protocolo OAI-PMH como meio de expor metadados para serem coletados por provedores de serviço ou agregadores, e os provedores de serviço (PS) usam metadados coletados automaticamente dos provedores de dados, via o Protocolo OAI-PMH, como base para oferecer produtos e serviços de valor agregado".

Esta Federação identifica como provedores de dados parceiros institucionais e/ou individuais que administrem repositórios em uma ou mais área de ciências da comunicação, com um ou mais tipo e suporte de fontes, e que se dispõem a adotar o protocolo OAI-PMH para intercâmbio de seus metadados. O papel de provedor de serviço é exercido, inicialmente, pela Portcom/Intercom, que assume a responsabilidade de desenvolver e implementar o coletador dos metadados dos parceiros/provedores de dados e oferecer interface unificada de busca.

Visando a oferecer suporte metodológico, procedimentos de trabalho e tecnologias específicas para aquelas instituições interessadas em implementar bibliotecas digitais com coleções digitais diversificadas (tipo de fonte ou de temática) e com serviços alinhados as dimensões "cognitivo

tático" e "cognitivo estratégico", a Federação projetou e desenvolveu vários módulos.

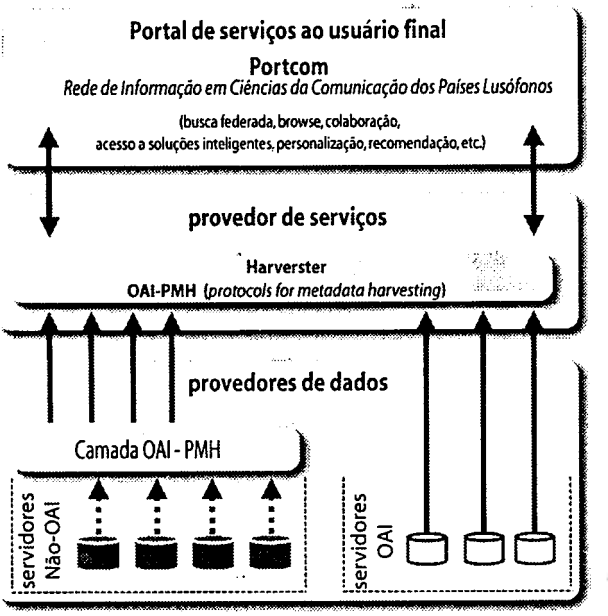


Figura 3 – Modelo operacional da Federação de Bibliotecas Digitais Lusófonas em Ciências da Comunicação

Módulo Revcom - Coleção eletrônica de revistas científicas lusófonas[5]

A Revcom abrange uma coleção selecionada de periódicos científicos publicados em países de língua portuguesa e tem como objetivo contribuir para o desenvolvimento da pesquisa científica lusófona, por meio do aperfeiçoamento e da ampliação dos recursos de disseminação, publicação e avaliação dos seus resultados, fazendo uso intensivo da publicação eletrônica.

Este módulo pressupõe uma série de atividades estrategicamente definidas para sustentar e fortalecer a atuação da Federação e, em contrapartida dos parceiros, ante às dimensões "cognitivo tático" e "cognitivo estratégico". Tais atividades englobam o seguinte: fomentar e articular a discussão e elaboração de indicadores e critérios de qualidade para a publicação científica da área de ciências da comunicação entre toda a comunidade científica; repassar à comunidade científica um modelo capaz de subsidiar a melhoria do padrão editorial das publicações nacio-

nais; aumentar a visibilidade, a acessibilidade e a credibilidade nacional e internacional da publicação científica em ciências da comunicação dos países de língua portuguesa; incrementar o impacto da produção científica lusófona, atuando diretamente no processo de comunicação científica.

Iniciado em 2002 com uso da metodologia Scielo[6], hoje adota o Sistema Eletrônico de Editoração de Revistas (SEER), que foi traduzido e customizado pelo IBICT baseado no *software* desenvolvido pelo Public Knowledge Project (Open Journal Systems) da Universidade British Columbia (http://www.pkp.ubc.ca/ojs/). Este sistema tem como objetivo dar assistência aos editores científicos em cada uma das etapas do processo de edição dos periódicos científicos, desde a submissão e avaliação dos consultores até a publicação *on-line* e sua indexação. Os sistemas de gerenciamento do periódico podem ser definidos de acordo com suas próprias políticas de publicação (Arellano, 2004).

Neste módulo, os parceiros/provedores de dados são os editores científicos de revistas publicadas por associações de pesquisa e/ou programas de pós-graduação. Eles são convidados, motivados e instigados a utilizar o aplicativo SEER (customizado para a área de comunicação pela equipe da Portcom) para o gerenciamento de suas respectivas revistas.

Módulo Arena Científica

Arena Científica é um repositório de *e-prints*[7], inspirado em um dos modelos de negócio instituído pela *Open Archives Initiative*[8] para viabilizar um sistema de comunicação científica eletrônica baseado no auto-arquivamento de textos completos pelos autores de forma a garantir o acesso livre e permanente em meio eletrônico e espelhar a produção científica de uma ou mais instituições e/ou áreas da comunicação.

Este módulo, fortemente embasado nas dimensões cognitivas definidas para a Federação, tem como proposta "ser um espaço de construção, reconstrução, compartilhamento e distribuição de conhecimento e inteligência visando a otimizar a colaboração e comunicação entre pesquisadores lusófonos da área das ciências da comunicação, incrementar o ciclo de geração de novos conhecimentos e promover acesso integrado à produção científica da área (Ferreira, 2002)[9].

Foi implementado com o *software* DICI (Diálogo Científico), versão brasileira e customizada pelo IBICT do *software Eprints Archive* desenvolvido

pelo Southampton University na Inglaterra (http://www.eprints.org). Este *software* promove um espaço virtual para o autodepósito de textos inéditos (papers, trabalhos de eventos, etc) ou já publicados (capítulos de livros, artigos de periódicos, etc), bem como para se pesquisar, comentar e/ou avaliar os textos depositados pelos pares.

O produto deste módulo está disponível no próprio Portal da Portcom para uso dos estudantes de pós-graduação, docentes e pesquisadores (considerados produtores de conteúdo e também usuários de informação científica) da área de ciências da comunicação.

Módulo Reposcom - Repositórios Institucionais em Ciências da Comunicação

O Reposcom busca oferecer um mecanismo fácil e ágil para o armazenamento, a divulgação e o acesso à produção intelectual em texto completo (formato estruturado ou não estruturado) de instituições. A implementação de repositórios locais, regionais e/ou nacionais contribui para o aumento da visibilidade e impacto desta produção, além de garantir a preservação da memória científica da área.

Este módulo focaliza a produção técnica científica do corpo docente, discente e pesquisadores de instituições parceiras, seja ela composta de publicações tradicionais como não tradicionais (monografias, livros, capítulos de livros, dissertações e teses, relatórios, artigos de periódicos, filmes e vídeos entre outros). Foi desenvolvido com base no *software* Dspace (http://www.dspace.org/faqs/#what), sistema de fonte aberta desenvolvido pelo Massachusetts Institute of Technology (MIT) e pela Hewlett-Packard (HP) que foi traduzido para o português e customizado pela própria equipe da Portcom.

O produto deste módulo está disponível para que os parceiros interessados (bibliotecas, núcleos de pesquisa, programas de pós-graduação, associações dentre outros) implementem repositórios para gerenciar a produção científica dos seus associados ou participantes.

Módulo Provedor de Serviços - Harvester

Módulo em fase de desenvolvimento e implementação. Seu propósito é desenvolver e implementar automaticamente um coletador de metadados para servir como provedor de serviços na Federação.

Este provedor está sendo desenvolvido de acordo com normas e padrões de normalização e descrição compatíveis com o protocolo Dublin Core, visando à oferta de mecanismo de busca integrada por meio de interface única com oferta de serviços de valor agregado ao usuários parceiros e/ou finais da Federação (serviço de News, Boletins de Alerta, Serviços de Personalização, Notas em Documentos, dentre outros). Como estratégia de implementação, o coletador está sendo pensando para a busca integrada dos conteúdos disponíveis nos *softwares*: Dspace; Seer, OJS, DICI e Eprints. Posteriormente, será iniciada a integração com provedores de dados/parceiros com plataformas distintas, mas que implementem o protocolo OAI-PMH, conforme definição da Federação.

Neste momento, estão sendo analisados os sistemas PKP / Public Knowledge Project (Open Journal Systems) da Universidade British Columbia (http://www.pkp.ubc.ca/ojs) e o ARC – A Cross Archive Search Service (http://arc.cs.odu.edu) para futura integração e implementação do provedor de serviço.

Desenvolvimento de estratégias de implementação da Federação

A implementação da Federação, como um todo, tem sido feita com a articulação de diversificadas ações, tais como:

- identificação, sensibilização e formalização de parcerias com diversificadas instituições lusófonas de ensino, pesquisa e/ou profissionais;
- definição conjunta quanto a forma de participação dos provedores de dados/parceiros;
- implantação dos módulos de provedores de dados e provedor de serviço: concepção, metodologia, tecnologias, treinamento e capacitação de Recursos Humanos, testes e pré-testes;
- monitoramento, análise de resultados e correção de rumos.

Para a implementação das Bibliotecas Digitais Federadas, pelos provedores de dados/parceiros, foram definidas as seguintes estratégias específicas:

- os provedores de dados/parceiros poderão implementar um ou mais módulos propostos pela Federação, ou participar apenas da coleta

automática de metadados por meio da instalação do protocolo OAI em seus respectivos ambientes;

• a equipe da Portcom assume a gestão das coleções parceiras para os Módulos Revcom e Reposcom, em regime de "incubação", durante período experimental;

• especificamente para o módulo Revcom, os critérios de qualidade de avaliação das revistas científicas em ciências da comunicação serão adaptados dos critérios Qualis da Capes;

• ainda em relação ao módulo Revcom, a equipe da Portcom assume a conversão para suporte eletrônico da coleção retrospectiva de 2001 a 2004 das revistas brasileiras classificação Nível Nacional A no Sistema Qualis da Capes;

• referente ao módulo Arena, a equipe da Portcom tem buscado fomentar parcerias com membros de Núcleos de Pesquisas, em especial dos Núcleos da Intercom, para a geração colaborativa de conhecimento científico utilizando este espaço como ambiente facilitador e integrador de suas atividades de pesquisa e produção.

• para todos os módulos foram preparados manuais de procedimentos e esquema para auxiliar na capacitação e treinamento de equipes dos provedores de dados/parceiros.

A articulação da Federação com a Portcom e Intercom é parte importante da estratégia de implementação, tendo em vista que, neste momento, esta é a executora, fomentadora da Federação bem como a gerenciadora do portal que dá acesso a todo estes serviços e atividades.

Até o momento, já foram formalizadas parcerias com seis revistas brasileiras e duas portuguesas. As brasileiras são as seguintes: *Revista Brasileira de Ciências da Comunicação da Intercom*; *Revista Comunicação & Sociedade* da Universidade Metodista de São Bernardo do Campo; Revista *Contracampo* da Universidade Federal Fluminense; *Revista Famecos* da Pontifícia Universidade Católica do Rio Grande do Sul e a *Revista Galáxia* da Pontifícia Universidade Católica de São Paulo. As revistas portuguesas são: *Revista Comunicação & Sociedade* da Universidade do Minho e *Media e Jornalismo* do Centro de Investigação Media e Jornalismo.

Atualmente estão "incubados", no Portal da Portcom, repositórios institucionais de todos os 22 núcleos de pesquisas da Intercom e ainda

do Núcleo de Comunicação Organizacional da ECA/USP. Inicia-se agora a implementação de repositórios institucionais pelos representantes regionais da Portcom[11] em suas respectivas instituições de origem visando a atuar como comunidades-piloto em regime experimental.

Gestão da Federação

A gestão das atividades e módulos propostos pela Federação adota o modelo de gestão compartilhada, no qual cada parceiro mantém sua identidade institucional e programática, valoriza-se o espírito de cooperação e co-responsabilidade, privilegia-se a ausência de hierarquia, a complementaridade e interdependência em prol do interesse do todo.

Tais valores levaram à definição de uma estrutura organizacional para a Federação em Ciências da Comunicação constituída por comitê consultivo, comitê operacional e secretaria executiva.

- Comitê Consultivo – define e valida política e estrategicamente as decisões da Federação de Bibliotecas Digitais Lusófonas em Ciências da Comunicação, estabelece padrões e critérios de funcionamento, analisa e aprova o plano de trabalho, coordena a divisão de trabalho entre as instituições, identifica oportunidades e *gaps* informacionais, propõem indicadores de desempenho dos produtos e serviços oferecidos e articula fontes de financiamento para projetos específicos. É composto por representantes de instituições fortemente comprometidas com o desenvolvimento do ensino, pesquisa e extensão na área de ciências da comunicação[10].
- Comitê Operacional – implementa e coordena bibliotecas digitais locais, regionais e/ou nacionais nas instituições parceiras a partir da metodologia da Federação, opera as fontes de informação descentralizadamente, mantém os conteúdos atualizados, participa de reuniões e capacitações técnicas com os demais parceiros para definir estratégias e acompanhar os ajustes e desenvolvimentos metodológicos e/ou tecnológicos. A participação neste Comitê é aberta às instituições lusófonas de ensino e pesquisa, desde que manifestem voluntariamente ou por convite o interesse em participar como provedora de dados/parceira e aderir aos padrões e modelos de qualidade estabelecidos.
- Secretaria Executiva – garante o pleno funcionamento da Biblioteca Digital Federada, coordena e promove a Federação, gerencia o funcio-

namento efetivo do Comitê Consultivo organizando reuniões periódicas, mantém a integridade e constante avaliação dos trabalhos, desenvolve padrões e critérios de qualidade, capacita equipes, assegura a atualização constante das ferramentas e produtos, coordena e atualiza constantemente a ferramenta de busca que integra os conteúdos das bibliotecas digitais federadas etc. Neste momento, a equipe da Portcom/Intercom (composta por bibliotecários, analistas e estagiários e ainda por seus representantes regionais) é quem assume a Secretaria Executiva da Federação.

O espaço de encontro entre as comunidades de ambas, as áreas ciência da informação e ciências da comunicação, para discussão e troca de experiências sobre as demandas e ofertas e para as reuniões dos Comitês desta Federação é mantido sistemática e anualmente durante os Congressos Brasileiros promovidos pela Intercom. Neste evento (que conta anualmente com participação de 3 a 4 mil pessoas) são realizadas reuniões de trabalho, oficinas para treinamentos de novos parceiros e módulos e, principalmente, o Endocom / Encontro de Informação em Ciências da Comunicação.

Quadro 1 – Temas discutidos no Endocom - 2000 a 2005

Ano	Evento/Local	Discussão em pauta ou módulo em pauta	Principal Palestrante
2000	X Endocom /UNAm -Manaus-AM	Multiculturalismo e a sociedade da informação.	Dr. *Eduardo Tadao Takahashi* – Coordenador do Programa da Sociedade da Informação, MCT, Brasil Dr. *Raul Fuentes Navarro*, coordenador da Comnet-AL e Professor do ITESO, México.
2001	XI Endocom Uniderp- Campo Grande - Mato Grosso do Sul	Diagnóstico da informação brasileira na área de Comunicação. *Lançamento do Portal* do PORTcom	Dra. *Sueli Mara S.P. Ferreira* – Coordenadora da Portcom/Intercom e Professora da USP, Brasil.
2002	XII Endocom Uneb – Salvador - Bahia	Diagnóstico da Informação brasileira na área de Comunicação. *Lançamento do módulo Arena*	Dr. *Eloy Garcia* – Diretor do IBICT, Brasil
2003	XIII Endocom PUC – Belo Horizonte – MG	Panorama Mundial das Revistas Científicas *Lançamento do módulo REVcom*	Dr. *Daniel Jones* – Universidade de Barcelona, Espanha
2004	XIV Endocom PUC – Porto Alegre – RS	A biblioteca digital fortalecendo a comunicação, os acontecimentos e a memória lusófona em ciências da comunicação *Lançamento do módulo REPoscom*	MSc *Eloy Rodrigues* – Universidade do Minho, Portugal
2005	XV Endocom UERJ – Rio de Janeiro - RJ	Discussão sobre o Movimento do Acesso Livre *Formalização da Federação de Bibliotecas Digitais*	Dr *Jean Claude Guedón* – University of Toronto, Canadá.

Nos últimos anos, o Endocom tem sido um dos pontos-chave na estratégia e gestão da Portcom e seus produtos, tendo em vista o ambiente propício que se tem criado para a troca de experiências, revisão e validação sistemática e periódica dos serviços e produtos e divulgação dentre a comunidade de novas tendências e inovações. Ano a ano cresce o número de pesquisadores, alunos e profissionais que participam, apresentam seus trabalhos e utilizam o Encontro para trocar experiências e fortalecer as redes de relacionamento. Os temas discutidos nos encontros desde o início do projeto da Federação estão indicados no quadro 1.

4. Próximos passos

Com o modelo da Federação definido e as várias metodologias e tecnologias já construídas, a Federação passa a focar-se no fomento, articulação e implementação de bibliotecas digitais federadas lusófonas (provedores de dados) em instituições de ensino, pesquisa e extensão em ciências da comunicação.

A ampliação desses provedores de dados/parceiros fortalece a necessidade de realização de estudos de usos e de acompanhamento, análise de resultados dos módulos desenvolvidos e implantados, assim como de enriquecimento de conteúdo, serviços e tecnologia.

O acompanhamento da evolução tecnológica e a análise de pertinência de novos *software* e aplicativos também compõem o quadro de ações que devem ser continuamente realizadas para garantir que a Federação mantenha a eficiência e a sustentabilidade de sua operação.

A Federação também desenvolverá e integrará novos conteúdos e serviços, em conjunto com seus parceiros e com o foco nas dimensões "cognitivo tático" e "cognitivo estratégico" do modelo adotado.

Igualmente importante e necessário na evolução das atividades da federação é o desenvolvimento de uma ferramenta de indexação, como tesauro e/ou vocabulário controlado para a área de ciências da comunicação, em português (atendendo a todas as especificidades dos países lusófonos).

Além desses temas, a preservação digital, direitos autorais e acessibilidade também estão na pauta de discussão e de execução e deverão ser tratados nas próximas etapas de desenvolvimento da Federação.

Notas

[1] Budapest Open Access Initiative URL: http://www.soros.org/openaccess/read.shtml

[2] Por esta ocasião, a Unesco implantou redes de informação em várias áreas do conhecimento. Em ciências da comunicação, além da Portcom, surgiram a Rede COMNET (internacional) e a COMNET-AL (América Latina), sendo que estas duas últimas foram extintas em 1998.

[3] Pesquisa desenvolvida em 2000 com uma equipe formada pelos representantes regionais da rede Portcom, sob a coordenação da professora doutora Sueli Mara S. P. Ferreira. Representantes: Elizabeth Heitor Pinto da Uam (região Norte); Maria das Graças Targino Moreira Guedes da UFPi (região Nordeste); Maria Auxiliadora Andrade Echegaray da UFG (região Centro-Oeste); Ilce Gonçalves Milet Cavalcanti do IBICT/UFRJ (região Sudeste); Ida Regina Chitto Stumpf, da UFRGS (região Sul) e Normanda Miranda Kiyotani da ECA/USP.

[4] O objetivo da interoperabilidade é construir serviços de forma coerente para os usuários, a partir de componentes que são distintos técnica e administrativamente em diferentes organizações (ARMS, 2002).

[5] URL: http://revcom.portcom.intercom.org.br

[6] Diante das dificuldades de customização e adequação da metodologia SCIELO/BIREME às especificidades da área de comunicação, optou-se por estacionar temporariamente as atividades de marcação do conteúdo e iniciar uma atuação diretamente no fluxo editorial, descentralizando o processo de tratamento dos conteúdos.

[7] Warner (2003, p. 2) esclarece que, segundo Paul Ginsparg, *e-print* é um trocadilho do termo *preprint* – em português pré-publicações. Mas o termo apareceu pela primeira vez na página '*Algebraic Geometry E-Prints*', criada por Dave Morrison, da Duke University, em fevereiro de 1992. Morrison, por sua vez, atribui a alcunha à Greg Lawler.

[8] Open Archives Iniciative – A Iniciativa de Arquivos Abertos (Open Archives Initiative - OAI) desenvolve e promove padrões de interoperabilidade para facilitar a disseminação de conteúdo científico. Surgiu em outubro de 1999, em Santa Fé, New México. Mais detalhes http://www.openarchives.org/documents/FAQ.html

[9] Representantes Regionais do Portcom: identificar

[10] Atualmente os membros deste Comitê são representantes das seguintes organizações: (a) Intercom/Sociedade Brasileira de Estudos Interdisciplinares em Comunicação; (b) Lusocom/Federação Lusófona de Ciências da Comunicação; (c) Representante da área de Ciências Sociais Aplicadas da Capes; (d) Representante da área de Comunicação junto

ao CNPq; (e) Programa de Apoio a Publicações Científicas da Fapesp; (f) Compos – Associação dos Programas de Pós Graduação em Ciências da Comunicação; (g) Abecom – Associação Brasileira de Escolas de Comunicação; (h) CBBU – Comissão Brasileira de Bibliotecas Universitárias.

[11] Os representantes regionais atuais da Portcom: Região Sul – professora doutora Patrícia Marchiori/UFPr; Região Sudeste – professora Eloisa da Conceicão Principe de Oliveira/UFRJ/IBICT; Região Centro-Oeste – Miguel Ângelo Arellano/ IBICT; Região Nordeste – professor doutor Marcos Galindo; Região Norte – professora doutora Célia Regina Simonetti/UFAm.

Referências

ARELLANO, M.(2004) Publicação científica e arquivos abertos. Palestra apresentada no SNBU, Natal. URL:http://www.bczm.ufrn.br/snbu2004/segundas/powerpoint/DICI-SEER_Out2004.ppt

ARMS, W.Y.(2001) Digital Libraries. MIT Press.

BUDAPESTE Open Acess Initiative (s.d.) URL://http://www.soros.org/openaccess/read.shtml) Acessado em Dezembro, 2004.

COLEMAN, A.; OXNAM, M. (2002) Interactional Digital Libraries: introduction to a special issue on Interactivity in Digital Libraries. Journal of Digital Information, v.2, n.4, 2002.

FENG, L.; JEUSFELD, M.; HOPPENBROUWERS, J. (2001).Towards Knowledge-Based Digital Libraries. SIGMOD Record, v.30, n.1 p. 41-46, Mar.

FENG, L.; JEUSFELD, M.; HOPPENBROUWERS, J. (2002) A Strategic Level for Scientific Digital Libraries. Proceedings 5th International Conference on Asian Digital Libraries: Digital Libraries: People, Knowledge, and Technology, p. 495-506.

FENG, L.; JEUSFELD, M.; HOPPENBROUWERS, J.(2005) Beyond information searching and browsing: acquiring knowledge from digital libraries. Information Process Management, v.1, n. 1, p. 97-120, Jan.

FERREIRA, S.M.S.P. (2002a). Portal em Comunicações: parâmetros para sua construção com base em diagnóstico da informação brasileira da área. In: Congreso Latino Americano da Ciencias de la Comunicación, 6., no GT Comunicación, Tecnología y Desarrollo, junio 5 a 8, 2002. Santa Cruz de La Sierra, Bolivia. Anais... São Paulo: ALAIC.

FERREIRA, S.M.S. P. (2002b). Arena científica: arquivos abertos em Ciências da Comunicação. São Paulo: ECA/USP. (Projeto desenvolvido em Parceria com o IBICT/ Programa Open Archives. Versão de novembro de 2002).

GONÇALVES, M.; FRANCE, R.K.; FOX, E. (2001). MARIAN: Flexible Interoperability for Federated Digital Libraries. Computer Science, v. 2163,

Proceedings of the 5th European Conference on Research and Advanced Technology for Digital Libraries, p. 173 - 186.

IBICT (s.d). Glossário BDTD. URL: http://bdtd.ibict.br/bdtd/glossario/glossario.jsp. Acessado em 20 de março de 2005.

LIU, Xiaoming (2002). Federating heterogeneous digital libraries by metadata harvesting.

LUSENET, Y. (2002). Digital heritage for the future. Cadernos BAD, v. 2, p. 15-27.

PIRRI,M.; PETTENATI, M.C.; GIULI, Dino (2002). Design of a Federation Service for Digital Libraries:the Case of Historical Archives in the PORTA EUROPA Portal (PEP) Pilot Project Proc. Int. Conf. on Dublin Core and Metadata for e-Communities. University of Florence.

WARNER, S.L. (2003). Eprints and the open archives initiatie. Library Hi Tech, v. 21, n. 2, p. 151-158.

As autoras agradecem a colaboração de Simone Weitzel, Silvia Francisca L. Gonçalvez e Roberto Seiti Yamashiro, respectivamente pesquisadora, bibliotecária e bolsista de iniciação científica, membros da equipe da Portcom para o desenvolvimento da Federação.

Este livro foi publicado no formato 17x24cm
Fontes *GoudySans LT BT*, *GoudySans MD BT* e *Arcade Acentuada*
Miolo em papel 75 g/m²
Tiragem: 200 exemplares
Impresso no Setor de Reprografia da EDUFBA
Impressão de capa e acabamento: ESB Serviços Gráficos

LIDIA B. TOUTAIN

Doutora em Filosofia pela Universidad Leon (Espanha). Mestre em Biblioteconomia.
Atua na área da informação, tendo dirigido e coordenado diversos projetos na área da C&T, na Secretaria de Planejamento, Ciência e Tecnologia do Estado da Bahia. Foi diretora do Sistema Estadual de Bibliotecas Públicas da Bahia. Coordenou o Mestrado Interinstitucional em Ciência da Informação oferecido pela Unb por meio de convênio com a UFBA. Primeira coordenadora do Curso de Mestrado em Informação Estratégica da EBD/UFBA. Atualmente é professora e Vice-Coordenadora do Programa de Pós-Graduação em Ciência da Informação no ICI/UFBA.

LUÍS SAYÃO

Físico, Mestre e Doutor em Ciência da Informação.
Chefe do CIN – Centro de Informações Nucleares da Comissão Nacional de Energia Nuclear (CNEN).
Representante do Brasil no INIS – International Nuclear Information System, Agência Internacional de Energia Atômica.
Cordenador da Red Regional de Información en el Area Nuclear (RRIAN).
Membro da Câmara Técnica do Documento Eletrônico (CTDE) do Conselho Nacional de Arquivos (CONARQ).

Caso pudéssemos viajar no tempo, possivelmente veríamos pessoas discutindo acaloradamente os impactos da imprensa de tipos móveis em suas atividades e funções. Aproximadamente 500 anos depois somos desafiados, como os monges copistas de então, a desbravar e adaptar nossas práticas a uma "nova tecnologia", que se coloca a serviço da memória, da informação e do entretenimento.

A velocidade das mudanças sociais, as dimensões simbólicas da informação e a crescente individualização dos mecanismos de comunicação humana tornam esta tarefa ainda mais complexa. Disponibilizar conteúdo em redes eletrônicas não é apenas romper barreiras de tempo e espaço. Tampouco é, por si só, uma quebra de paradigma nas condições de apreensão de conhecimentos. Uma sociedade da informação apenas se configurará, de fato, quando as exclusões físicas e intelectuais de acesso forem superadas; quando a produção, consumo e disponibilidade de conteúdos informativos forem sustentados por critérios transparentes de qualidade; e quando houver efetivo comprometimento ético quanto ao uso de expressões intelectuais produzidas por terceiros. Tais condições – entre outras – devem necessariamente compor o pano de fundo das ações relativas à criação de bibliotecas digitais.

A Universidade Federal da Bahia (UFBA) e o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) oferecem um espaço extremamente oportuno para que prestigiados especialistas destaquem facetas contextuais, tecnológicas, de gestão e uso destes ambientes, além da descrição de experiências de significativa relevância para os interessados no assunto.

Considerando a dinâmica do tema e a carência de textos em língua portuguesa, espera-se que esta coletânea estimule outras iniciativas que, em conjunto, constituam amplo repertório teórico-prático para a comunidade de profissionais da informação.

Profª. Dra. Patricia Zeni Marchiori
UFPR/SCSA/DECiGI
pzeni@ufpr.br

ISBN 85-232-0350-8
9 788523 203504