

O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais

The challenge of the interoperability and the new perspectives for digital libraries

Luis Fernando SAYÃO¹

Carlos Henrique MARCONDES²

RESUMO

O acesso integrado e transparente a recursos informacionais heterogêneos, armazenados em bibliotecas digitais e repositórios digitais distribuídos e gerenciados de forma autônoma por instituições diferentes, é um dos grandes desafios das pesquisas na área de bibliotecas digitais. Dentro dessa perspectiva, o trabalho revisa os principais problemas de interoperabilidade entre bibliotecas digitais: os tipos de soluções, os níveis de interoperabilidade, a arquitetura de sistemas federados, a questão dos direitos autorais e os parâmetros de avaliação. Considera também como as tecnologias da informação e a Web Semântica estão sendo usadas para desenvolver sistemas de bibliotecas digitais interoperáveis. Conclui relacionando as principais questões ainda em aberto e que estão, no momento, sendo objetos de pesquisa para implementação da próxima geração de bibliotecas digitais.

Palavras-chave: interoperabilidade; bibliotecas digitais; sistemas distribuídos; federação de repositórios digitais; gestão de direitos no ambiente digital.

ABSTRACT

The integrated and transparent access to heterogeneous information resources, stored in distributed digital library and repositories, autonomously managed by different institutions is an important challenge in the digital library research area. From this perspective, this work reviews the main problems to achieve interoperability between digital libraries: different technical solutions, levels of interoperability, federated system architecture, copyright issues and evaluation parameters of interoperability solutions. It analyses also how the new development in Information Technology and Semantic Web have been applied to improve interoperable digital library systems. As conclusions it discusses the main challenges facing present digital library research to implement the next generation of digital libraries.

Keywords: interoperability; digital libraries; distributed systems; digital repositories federation; digital rights management (DRM).

¹ Chefe do Centro de Informações Nucleares, Comissão Nacional de Energia Nuclear. Rua General Severiano, 90, Botafogo, 22294-900, Rio de Janeiro, RJ, Brasil. Correspondência para/Correspondence to: L.F. SAYÃO. E-mail: <lsayao@cnen.gov.br>.

² Docente, Instituto de Artes e Comunicação Social, Departamento de Documentação, Universidade Federal Fluminense. Rua Prof. Lara Vilela, 126, São Domingos, 24210-590, Niterói, RJ, Brasil. E-mail: <marcon@vm.uff.br>.
Recebido em 2/5/2008 e aceito para publicação em 31/10/2007.

INTRODUÇÃO

A informação é produzida de forma dispersa no espaço e no tempo, e as bibliotecas sempre foram, ao longo da História, lugares ou instituições que se opunham a essa dispersão, concentrando a informação num lugar físico para servir a uma determinada comunidade de usuários. Esse é o valor que agregam. Como, porém, as bibliotecas tinham limites físicos, paredes, coleções em estantes, etc., o alcance desses serviços ficava restrito aos membros da comunidade que conseguia ter acesso presencial a ela.

Com o surgimento da Internet, essa situação evoluiu de forma avassaladora. Nos nossos dias, mais e mais registros da cultura humana são produzidos já diretamente em formato digital como música, imagens, vídeos, material textual – inclusive aquele a ser impresso em papel –, e ainda as novas formulações e concepções de registros proporcionados pela tecnologia. A Internet é o retrato vivo da “explosão informacional”, do “caos informacional”, mas o acúmulo imenso e desordenado de informação é, ao mesmo tempo, a sua fortaleza e a sua fragilidade, sendo necessário, portanto, algum grau de ordenamento e intermediação.

O ser humano individualmente não tem capacidade cognitiva em absorver esse volume crescente de informação colocado à sua disposição. Portanto são necessários sistemas de intermediação que se interponham entre o usuário e as fontes de informação que agreguem valor ao avaliar/selecionar/filtrar a informação. Com a realização do conceito de biblioteca digital, além da capacidade de coletar e concentrar informações dispersas ter aumentado enormemente, acrescentou-se ainda – apoiado fortemente em tecnologia da informação – algo de crucial importância: o potencial de atender a uma comunidade, que não se restringe mais a quem tem acesso presencial à biblioteca. Além disso, abriu-se a possibilidade inédita de um número ilimitado de usuários poder acessar simultaneamente a mesma cópia de um documento digital.

Essas características das bibliotecas digitais ampliam de forma extraordinária as suas possibilidades, tomando-as um instrumento de grande efetividade para distribuição, cooperação e acesso ao conhecimento, atendendo e podendo servir de pólo agregador para comunidades segmentadas e distribuídas geograficamente. Isso, se devidamente explorado, pode

representar um impacto importante para um país como o Brasil, que se caracteriza pela diversidade em todas as suas dimensões. Pode-se resumir que a principal vantagem das bibliotecas digitais sobre as bibliotecas físicas é a capacidade que elas têm de multiplicar o alcance – geográfico e temporal - em termos das comunidades que elas são capazes de atingir e servir.

Como ilustração, imagine o seguinte caso de uso: como integrar numa biblioteca digital conteúdos em cultura brasileira e língua portuguesa sobre “o Brasil colônia e a influência da cultura negra”, em formato digital disponibilizados na Internet, para fins de apoiar atividades educacionais de professores brasileiros? Conteúdos de interesse para essa proposta vão desde coleções custodiadas por museus, arquivos e bibliotecas, em diferentes graus de tratamento técnico informatizado, usando nenhuma ou diferentes tecnologias, padrões, protocolos e estágios de interoperabilidade, até páginas web simples de grupos Afro, ONGs, grupos de teatro e programas governamentais. Esses diferentes conteúdos só poderão ser integrados e reusados, no sentido de terem aproveitadas as sinergias uns dos outros, se estiverem ancorados por sistemas que permitam um alto grau de interoperabilidade.

Nos últimos anos, a interoperabilidade tem sido um dos itens mais críticos para quem pensa no desenvolvimento e operação de sistemas de repositórios e de bibliotecas digitais distribuídos funcionando em rede. O conceito de interoperabilidade, entretanto, está longe de ser uma novidade no domínio das bibliotecas: há muito tempo, desde meados do século XX, para fazer frente ao fenômeno social da “explosão informacional”, as bibliotecas sempre estabeleceram serviços cooperativos, trocaram informações, criaram um ordenamento universal dessas informações. Toda uma estrutura global foi montada em torno da idéia do compartilhamento e da cooperação entre bibliotecas.

Com a concretização e a consolidação do conceito de bibliotecas digitais – que se localiza na interseção entre biblioteconomia, ciência da computação e tecnologias de rede – a interoperabilidade torna-se um foco de grande interesse para muitos atores. Uma razão para que ela tenha recebido tanta atenção é que o problema permeia quase todos os aspectos circunscritos pela idéia de biblioteca digital, enquanto sistema implementado segundo uma arquitetura distribuída. Outro dado relevante é o crescente interesse da indústria de conteúdos, principalmente por parte

dos setores ligados à editoração científica, de incorporar nos seus modelos de negócio as novas formas de disseminação dos repositórios digitais como meio de distribuição de seus produtos no ambiente de rede. Isso se desenrola no contexto de uma economia da informação em transição, que está saindo de um modelo baseado em assinaturas para modelos baseados no acesso e uso. O acesso via *pay-per-view* a um artigo científico serve como exemplo de uma modalidade que se está popularizando. Nesse escopo está incluído ainda o interesse pelas tecnologias automatizadas de controle de direitos autorais como parte integrante das estratégias de interoperabilidade.

Há ainda o interesse de instâncias governamentais nas potencialidades dos novos serviços que se tornam possíveis com a integração de acervos digitais heterogêneos e distribuídos. Uma das principais áreas onde a integração e o compartilhamento são altamente demandados é a da educação, particularmente no ensino a distância. A adoção generalizada das tecnologias Internet como canal para a educação e treinamento tem resultado numa abundância de recursos educacionais e de treinamento em formato digital pronto para utilização via web. Não obstante a sua aparente onipresença, a localização e reuso desses objetos educacionais são prejudicados pela falta de esforços coordenados e políticas voltadas para o armazenamento, o tratamento técnico e a gestão de direitos visando a uma integração enriquecedora desses conteúdos (Hatala et al., 2004).

Em paralelo às bibliotecas, os arquivos e ainda os museus têm também vislumbrado a Internet como um mecanismo por meio do qual é possível ampliar de uma forma nunca antes possível as fronteiras físicas do mundo convencional, criando novas possibilidades de acesso e disseminação de suas coleções. Em relação a esse tema, pode ser interessante consultar Warren, Thurlow, Alsmeyer. (2006), Projeto DigiCult (<http://www.digicult.info/pages/info.php>) e página "Archives & Museums Informatics", (<http://www.archimuse.com>).

Nesse contexto, tendo a percepção mais imediata das vantagens dos repositórios digitais, dispendo de alguns recursos financeiros, humanos e metodológicos e tendo ferramentas de *software*, via de regra, livremente disponíveis, as organizações da área de conhecimento – principalmente as bibliotecas, arquivos e museus – têm crescentemente migrado seus estoques de informação para repositórios digitais. Caracte-

risticamente esses repositórios são estanques e fragmentados e as buscas a eles devem ser realizadas individualmente, por meio de interfaces e formulações distintas. Quem desejar integrar dados de diferentes locais terá que fazê-lo manualmente. Esse enfoque tem produzido muitos sistemas de repositórios digitais heterogêneos, mas torna a interoperabilidade, o reuso, o intercâmbio e o desenvolvimento cooperativo extremamente difíceis de se alcançar. Isso não é exatamente o que se espera como resultado dessa transição para um novo conceito de integração de bibliotecas.

O acompanhamento do estado-da-arte na área de bibliotecas digitais tem mostrado como evoluiu o enfoque da integração de dados nesse domínio: no primeiro momento caminhou-se no sentido de se criar bibliotecas digitais isoladas. Nesse patamar, as bibliotecas são mantidas por uma única instituição e as coleções de dados são autocontidas, enquanto os conteúdos são formalmente localizados e gerenciados de forma centralizada. O passo seguinte está relacionado à federação em rede de várias bibliotecas independentes, possivelmente organizadas em torno de um tema ou área comum, formando uma rede de bibliotecas acessível por meio de uma única interface (Pirri; Pettenati; Giuli, 2002).

Existem diferentes modelos, tecnologias e metodologias para se alcançar o nível necessário de interoperabilidade entre bibliotecas digitais e seus componentes de serviço, tendo como perspectiva as razões e os fundamentos do sistema de informação como um todo. É esse tema, precisamente, que vai ser discutido aqui.

ANTECEDENTES – O PROBLEMA

Há um consenso absoluto de que a interoperabilidade é um problema de vasto domínio. Contudo, ela tem sido investigada tipicamente dentro de escopos específicos, tais como os circunscritos por uma comunidade particular, por exemplo: bibliotecas, empresas, comunidades científicas; dentro de uma classificação particular de informação, por exemplo, registros eletrônicos, relatórios técnicos, *software*; ou ainda dentro de área particular de tecnologia da informação (TI), como bases de dados relacionais, imagens digitais, visualização de dados. Escapando desse fracionamento, as pesquisas mais recentes sobre

interoperabilidade no âmbito da arquitetura de bibliotecas digitais estão concentradas no desafio de *criar uma infra-estrutura para acesso e integração de informação transversalmente a esses domínios específicos*. Um objetivo comum desses esforços é permitir que diferentes comunidades, com diferentes tipos de informação e usando diferentes tecnologias, consigam um nível geral de compartilhamento de informação e, por meio de processos de agregação apoiados por tecnologia da informação, criem novos e mais poderosos tipos de serviços de informação (Payette et al., 1999).

Idealmente, uma biblioteca digital deve ser capaz de armazenar uma variedade de tipos tradicionais de conteúdo – livros, periódicos, relatórios técnicos, *softwares* –, bem como entidades multimídia complexas que misturam texto, imagens, vídeo e dados. Para que o acesso a essas informações seja efetivamente viável, o sistema no qual elas estão armazenadas deve ser capaz de gerar processos que sejam interoperáveis com os sistemas que estão à sua volta. Uma organização verdadeiramente interoperável é capaz de maximizar o valor e o potencial de reuso da informação que está sob o seu controle. É também capaz de intercambiar efetivamente estas informações com outras organizações igualmente interoperáveis, permitindo que novos conhecimentos possam ser gerados a partir da identificação de relacionamentos entre conjuntos de dados previamente não relacionados. Na perspectiva do usuário, as interfaces devem apresentar para o usuário uma visão unificada em termos semânticos de diferentes recursos informacionais heterogêneos, ou seja: como nomeá-los, como referenciá-los, como utilizá-los em buscas, como acessá-los, como apresentá-los para o usuário.

O que se espera como resultado das pesquisas em bibliotecas digitais é que se possa estabelecer uma infra-estrutura conceitual, tecnológica, metodológica, de serviços e gerencial coerente que permita: que sejam sempre possível buscas simultâneas em coleções múltiplas, heterogêneas e operadas por sistemas tecnologicamente diferentes; que essas coleções possam estar custodiadas por organizações distintas; que possam estar geograficamente espalhadas por todo o mundo; que as buscas possam ser realizadas pelo usuário final de maneira fácil e transparente e utilizando ferramentas comuns de acesso à Internet, como são os navegadores *web*. Além do mais, essa infra-estrutura deve ser tal que possa oferecer ainda ferramentas para gestão de direitos e propriedade intelectual, e finalmente,

considerar a identidade do usuário e seus direitos individuais e institucionais de acesso.

O QUE É INTEROPERABILIDADE

O Online Dictionary for Library and Information Science (ODLIS), define o termo interoperabilidade como:

“A capacidade de um sistema de hardware ou de software de se comunicar e trabalhar efetivamente no intercâmbio de dados com um outro sistema, geralmente de tipo diferente, projetado e produzido por um fornecedor diferente.” (Online..., 2004).

Para a área de tecnologia da informação, há um consenso geral de que interoperabilidade é algo como “a capacidade de computadores e programas de fabricantes diferentes trocarem informações”. No contexto das bibliotecas, porém, o conceito de interoperabilidade não está circunscrito somente a uma questão de comunicação entre componentes de um sistema de computadores. Mais especificamente no âmbito das bibliotecas digitais, o conceito de interoperabilidade é complexo e estratificado, refletindo a diversidade de visões, o número de variáveis envolvidas e a interdisciplinaridade que está subjacente e ele. Por exemplo, no contexto do importante trabalho desenvolvido em colaboração pelo Digital Library Research Group da Cornell University e a Corporation for National Research Initiative (CNRI), interoperabilidade é definida como “a capacidade de componentes ou serviços de bibliotecas digitais serem funcionalmente e logicamente intercambiáveis em virtude deles terem sido implementados de acordo com um conjunto de interfaces bem definidas e publicamente conhecidas” (Payette, 1999, p.2). No modelo implementado, tendo como referência essa definição, serviços e componentes distintos podem comunicar-se mutuamente por meio de interfaces abertas, e os usuários podem interagir com eles de maneira equivalente.

Na visão da Ukoln (2005), expressa também por Miller (2000), a interoperabilidade pode ser considerada como o processo contínuo de assegurar

que sistemas, procedimentos e cultura de uma organização sejam gerenciados de tal forma que possibilitem a maximização das oportunidades para intercâmbio e reuso de informação.

A partir dessas definições, fica claro que a interoperabilidade está longe de depender somente de requisitos técnicos – como, por exemplo, o uso de programas e computadores compatíveis –, embora possa ser importante em algumas situações. Assegurar a plena interoperabilidade exige freqüentemente uma mudança profunda na forma pela qual uma biblioteca digital trabalha, relaciona-se com as organizações parceiras, usuários e fornecedores e, especialmente, sua atitude diante dos problemas relacionados à informação. O desafio de projetar serviços coerentes para uma diversidade de usuários a partir de componentes que são tecnicamente diferentes e gerenciados por diferentes organizações exige um sofisticado grau de cooperação que pode ser diferenciado em pelo menos três instâncias (Arms, 2000; Arms et al., 2002):

- a) acordos técnicos - cobrem formatos, protocolos, sistemas de segurança de forma que mensagens possam ser trocadas;
- b) acordos sobre conteúdos – cobrem dados e metadados e incluem acordos semânticos sobre interpretação das mensagens;
- c) acordos organizacionais – cobrem as regras básicas para acesso, para mudanças nas coleções e serviços, pagamento, autenticação, etc.

Soma-se ainda uma instância de acordos políticos, onde são estabelecidos os fóruns necessários e definidas as diretrizes e as políticas concernentes, incluindo as questões de financiamento.

AS MUITAS FACES DA INTEROPERABILIDADE

A face mais visível da interoperabilidade é certamente a interoperabilidade técnica, por ser, talvez, o aspecto mais perceptivelmente responsável por manter os sistemas de informação interoperáveis. A interoperabilidade tem muitas outras faces, entretanto, cada uma delas revelando um aspecto da complexidade de efetivamente se ter bibliotecas digitais funcionalmente integradas. Portanto outros aspectos não menos

relevantes devem ser também considerados (Miller, 2000; Ukoln, 2005):

- a) Interoperabilidade técnica – as considerações sobre os aspectos técnicos incluem assegurar envolvimento de um conjunto de organizações no contínuo desenvolvimento de padrões de comunicação, transporte, armazenamento e representação de informações, tais como são o Z39.50 (<http://www.loc.gov/z3950/agency/>), Search Retrieval Web Service (SRW) (<http://www.loc.gov/standards/sru/srw/index.html>), ISO-ILL e o XML (<http://www.w3.org/XML/>). Inclui também os esforços cooperativos para assegurar que padrões individuais evoluam em benefício da comunidade envolvida e para facilitar, onde for possível, convergência desses padrões, de forma que seja possível que os sistemas possam ter como base mais de um conjunto de padrões.
- b) Interoperabilidade semântica – está relacionada com o significado ou semântica das informações originadas de diferentes recursos e é solucionada pela adoção de ferramentas comuns ou/e mapeáveis de representação da informação, como esquemas de metadados, classificações, tesouros e mais recentemente, ontologias; um exemplo de questão endereçada por essa faceta da interoperabilidade pode ser o seguinte: o que significa “autor” para um recurso informacional? Será a mesma coisa que “criador” para um outro recurso?
- c) Interoperabilidade política/humana – independente das questões relacionadas à maneira pela qual a informação é descrita e disseminada, a decisão de tornar os recursos informacionais mais amplamente disponíveis e interoperáveis tem implicações para a organização, para as equipes envolvidas e para os usuários em termos comportamentais, de recursos e de treinamento. A ênfase dada por parte de alguns setores governamentais aos problemas de democratização do acesso, da exclusão digital e da federação de fontes de informação voltadas para a educação a distância, tem impacto nas políticas públicas para a área, e estão enquadrados nesse item.
- d) Interoperabilidade intercomunitária – enfoca a necessidade, cada vez mais urgente,

impulsionada pela crescente interdisciplinaridade, principalmente nas áreas de pesquisa, de acesso a informações provenientes de um espectro amplo de fontes distribuídas por organizações, áreas de conhecimento e comunidades de natureza distintas. Geralmente exige o estabelecimento de fóruns para discussão e consenso em torno de práticas e procedimentos comuns.

- e) Interoperabilidade legal – considera as exigências e as implicações legais de tornar livremente disponíveis itens de informação;
- f) Interoperabilidade internacional – quando se atua em escala internacional é necessário contornar a diversidade de padrões e normas, os problemas de comunicação, as barreiras lingüísticas, as diferenças no estilo de comunicação e na falta de uma fundamentação comum.

NIVEIS DE INTEROPERABILIDADE

O nível de interoperabilidade é referido aqui como o grau de compromisso ou acoplamento entre sistemas (instituições, bibliotecas digitais) para torná-los interoperáveis e seria uma medida do esforço para torna-se interoperável. Arms et al. (2002) relacionam as funcionalidades ou facilidades oferecidas aos usuários resultantes de um alto nível de interoperabilidade entre diversas bibliotecas digitais *versus* o custo de adesão a esse nível de funcionalidade por parte de novos parceiros: quanto maior o nível de interoperabilidade, maior o custo ou esforço para que novos parceiros adiram à iniciativa. Esses autores, no contexto do desenvolvimento da NSDL (*National SMETE Digital Library*), identificam três níveis de interoperabilidade aplicáveis ao domínio das bibliotecas digitais: *federação*, *harvesting* (colheita automática de metadados) e *gathering* (agregação automática de informação).

O nível mais alto, a federação, corresponde à mais robusta forma de interoperabilidade; em contrapartida, é a que exige maior esforço dos participantes. Para efetivar-se, ela exige que um grupo de organizações concorde que seus serviços estejam em conformidade com um conjunto de especificações,

geralmente selecionadas a partir de padrões formalizados. Os termos “heterogêneo” e “federado” são freqüentemente usados para descrever sistemas cooperativos nos quais componentes individuais são projetados ou operados de forma autônoma. Esse tipo de cooperação está em contraste com o termo geral “sistema distribuído”, que também inclui coleções de componentes desenvolvidos em diferentes sites, mas que são cuidadosamente projetados para trabalhar em conjunto (Paepcke et al., 1998).

O principal desafio que se coloca na formação de federações é o esforço despendido por cada organização em implementar e manter atualizados todos os níveis dos acordos. As bibliotecas que compartilham registros de catálogos *on-line* usando o protocolo Z39.50, trabalham segundo o nível de federação. O ANSI/NISO39.50 (ISO 23950) é um protocolo de comunicação entre computadores que pode ser implementado sobre qualquer plataforma. Ele tem como propósito a pesquisa e a recuperação de informações. A implementação do protocolo permite que, por meio de uma única interface, seja possível o acesso uniforme a uma diversidade de fontes de informações heterogêneas de modo síncrono e transparente para o usuário-final (NISO, 2002). Outro padrão nesta área é o Search Retrieval Web Service (SWR) (<http://www.loc.gov/standards/sru/srw/index.html>), um protocolo que se propõe a ser o sucessor do Z39.50.

As dificuldades de se criar grandes federações, porém, foi a principal motivação para busca de soluções menos onerosas para o estabelecimento de interoperabilidade entre bibliotecas digitais. Idéia subjacente é que os participantes concordem em despende um pequeno esforço que possibilite o compartilhamento de alguns serviços básicos, sem que seja necessário o enquadramento a um conjunto completo de acordos. Nessa situação enquadra-se o conceito de colheita automática de metadados (*metadata harvesting*), estabelecido pelo protocolo OAI-PMH (Open Archive Initiative Protocol of Metadata Harvesting). Enquanto os serviços baseados em *harvesting* são assíncronos e menos sofisticados do que os providos pelas federações, a sobrecarga sobre os participantes é consideravelmente menor. Como resultado, muito mais organizações, especialmente as surgidas no seio da academia, estão optando por esse tipo de interação, o que é provado pela rápida aceitação do OAI-PMH como um protocolo essencial nas

transformações que vêm ocorrendo nos padrões de comunicação científica (Marcondes; Sayão, 2001).

Ainda que um determinado grupo de organizações não estabeleça nenhum grau formal de cooperação, um nível básico de interoperabilidade é ainda possível por meio de agregação automática de informações disponíveis publicamente, utilizando-se metabuscadores, robôs, máquinas de busca e ainda por meio dos protocolos que suportam *web services* e outros padrões da indústria de TI. A agregação requer essencialmente pouco ou nenhum esforço por parte dos participantes, entretanto oferece um grau baixo de interoperabilidade (Arms, 2002).

Um nível ainda mais formalizado do que a federação pode ser ainda postulado: trata-se da padronização. Nesse nível, cada aspecto da interoperabilidade é formalmente definido e cada organização tem o rígido compromisso de seguir exatamente o conjunto de padrões e procedimentos convencionados. Na prática, isso pode determinar o uso da mesma plataforma computacional – *hardware*, aplicativos e sistema operacional – e das mesmas condicionantes administrativas, reduzindo drasticamente a autonomia dos componentes individuais. É esse nível que geralmente se estabelece no âmbito das redes cooperativas de bibliotecas que utilizam formatos de intercâmbio padronizados como Lilacs e MARC. É uma solução pouco provável além das fronteiras corporativas e de redes e sistemas altamente formalizados (Arms, 2002).

Na sua fase inicial, os projetos de bibliotecas digitais adotavam um único mecanismo de interoperabilidade, como por exemplo, o protocolo OAI-PMH, ou o protocolo Z39.50. É importante observar que as estratégias adotadas por sistemas importantes e de espectro amplo como o NSDL e como era o projeto original da Biblioteca Digital Brasileira (<http://bdtd.ibict.br/>) (Marcondes; Sayão, 2001), apontam para soluções em *diversos níveis simultâneos*, em que o grau de formalização das parcerias e o nível de interoperabilidade vão depender da importância das coleções, dos tipos de serviços e do grau de esforço necessário para operacionalizá-los. É possível, portanto, no contexto de um único sistema, estabelecerem-se federações distintas em torno de coleções especiais e de acordos técnicos e de conteúdo além de acordos organizacionais específicos (por exemplo, coleção de material didático), onde a

interoperabilidade menos formal – *harvesting* e *gathering* – também é considerada importante forma de cooperação. Ressalta-se, por último, que o termo *federação*, apesar de indicar um nível específico de interoperabilidade, tem sido muito freqüentemente usado para indicar genericamente a integração e a interoperabilidade entre repositórios digitais em diferentes níveis e operando simultaneamente, principalmente por autores mais próximos das áreas de TI.

TECNOLOGIAS DE INFORMAÇÃO E BIBLIOTECAS DIGITAIS

Ao contrário das metodologias usadas, num passado não muito distante, para automação de bibliotecas (Sayão; Marcondes, 1989), que tipicamente utilizavam tecnologias, ferramentas e padrões específicos da área bibliográfica, as bibliotecas digitais, numa visão otimista, podem incorporar quase que imediatamente e com poucos limites os progressos das Tecnologias de Informação e Comunicação (TICs), que fazem avançar quase que cotidianamente outras áreas importantes, como comércio eletrônico e telefonia móvel. Numa avaliação rápida sobre as áreas de pesquisas em bibliotecas digitais, principalmente no que concerne a tecnologias e arquiteturas, pode-se afirmar, sem muito risco, que elas são compostas da união de subáreas provenientes de vários domínios. Grande parte dos problemas específicos relacionados à construção de bibliotecas digitais têm sido enfocados como variações de problemas pertinentes a outros campos, como veremos nesta seção.

O acesso a recursos da Internet qualificados como distribuídos e heterogêneos é seguramente, como já enfatizado, um dos problemas críticos que devem ser superados na pesquisa e desenvolvimento da próxima geração de bibliotecas digitais. Os repositórios digitais disponíveis em rede, via de regra, variam em termos de interfaces, de representação de dados e, em geral, esses repositórios têm esquemas de metadados diferentes em termos de número de campos, de sintaxe e de semântica, e ainda são acessados por meio de protocolos incompatíveis e permeados por esquemas de proteção de propriedade intelectual diversificados. Tudo isso coloca o problema de desenvolver arquiteturas para federação de recursos heterogêneos, tendo como foco a tentativa de estabelecer pontes entre diferentes sistemas de informação e de representação, um dos principais

objetos de pesquisa na área. O objetivo de uma arquitetura de federação de serviços é oferecer uma interface uniforme para os recursos individuais, bem como uma visão integrada sobre os dados. Portanto a arquitetura deve ser concebida de forma a aceitar consultas sobre uma visão global das informações, baseada num modelo de dados uniforme.

O projeto de arquitetura pode ser focado de diversas formas, porém, de maneira genérica, a arquitetura de federação de serviços pode ser estruturada em três camadas distintas: 1) camada de repositórios digitais, onde as informações estão armazenadas com autonomia de representação e de interfaces de acesso; 2) camada de adaptação, que prevê acesso uniforme às informações ocultando as diferenças de modelos de dados e de interfaces de consulta. Nessa camada, adaptadores especiais ou mediadores – por exemplo, *harvesters* - têm que ser implementados para transformar os modelos específicos das fontes de dados em um modelo global do sistema federado. O mapeamento de esquemas particulares de metadados usados por cada repositório em um padrão comum, por exemplo, Dublin Core, serve como ilustração dessa camada; e 3) camada de federação, que responde pela integração global dos dados. Essa camada oferece os serviços para definição de uma visão integrada dos dados e consultas. É nessa instância que se pode dispor de bases de dados para descrever, por meio de metadados, os diferentes recursos disponíveis.

Muitos avanços importantes têm sido alcançados nos últimos anos, especialmente no desenvolvimento de mediadores que acessam as informações de fontes múltiplas. Um mediador tipicamente recebe uma requisição, por exemplo, uma consulta, e submete uma versão traduzida dessa requisição às várias fontes de informação, recupera e integra as respostas e apresenta ao usuário (Melnik; Garcia-Molina; Paepcke, 2000).

Na implementação de arquiteturas de federação, várias metodologias comuns à área de TI estão sendo utilizadas isoladamente ou em conjunto, ou são partes integrantes de projetos de pesquisa importante que estão em andamento. A seguir, relacionam-se algumas dessas metodologias:

Web Semântica - é uma extensão da Web atual -, que é formada por documentos compreensíveis unicamente por pessoas - para uma Web em que documentos seriam auto-descritíveis, de forma que seu conteúdo possa ser “compreendido” por programas

especiais, os agentes inteligentes de software (descrito mais adiante), que assim poderiam “raciocinar” e fazer inferências sobre o conteúdo de documentos, ajudando as pessoas em diferentes tarefas de recuperação e compartilhamento de informações que exijam raciocínio, decisões, inferência de conclusões a partir de informações não explicitamente disponíveis ou de informações contextuais. Nas palavras de Berners-Lee e seus colaboradores, a Web Semântica “é uma extensão da Web atual, na qual é dada à informação um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação”. (Berners-Lee; Hendler; Lassila, 2001). Há uma grande expectativa em torno das tecnologias da Web Semântica e as suas implicações para o futuro da cultura humana, que são destacadas pelos seus próprios criadores:

A Web Semântica não é meramente a ferramenta para administrar tarefas individuais que temos discutido até agora. Além do mais, se adequadamente projetada, a Web Semântica pode ajudar a evolução do conhecimento humano como um todo. (Berners-Lee; Hendler; Lassila, 2001).

Espera-se que tecnologias dotadas de habilidades semânticas possam ajudar os usuários de bibliotecas digitais de várias formas. As pesquisas nessa área, de forma geral, investigam como as tecnologias da Web Semântica podem potencializar as funcionalidades das bibliotecas digitais, especialmente na descoberta de recursos com mais eficiência e efetividade, e no compartilhamento de conhecimento no escopo da comunidade do usuário circunscrita pela biblioteca digital. Cenários típicos de uso de tecnologias semânticas em bibliotecas digitais incluem ainda, entre muitos outros usos, interfaces e interação homem-computador, exibição de informação, permissão para visualização e navegação em grandes coleções, construção de perfis de usuário, personalização e interação de usuários.

É consenso que uma das maiores fragilidades da Internet atual é que ela, apesar de ser adequada para a compreensão humana, não favorece a compreensão por máquina, de forma que a interpretação da informação contida em um dado documento necessita sempre da interferência humana. Para superar esses problemas, as ontologias

recentemente tornaram-se um foco de intenso interesse da Ciência da Computação. As ontologias estabelecem uma compreensão compartilhada de um domínio de interesse para apoiar a comunicação entre seres humanos e agentes computacionais. As ontologias são representadas caracteristicamente por uma linguagem de representação processada por computador, sendo considerada uma tecnologia-chave para o desenvolvimento da Web Semântica (Sure; Studer, 2005). Dessa forma, por meio do estabelecimento de esquemas comuns na forma de ontologias, as tecnologias semânticas permitem a descrição de objetos e repositórios digitais com o objetivo principal de capacitar a interoperabilidade, como será visto a seguir.

- *Ontologias* – a nova geração de sistemas de informação deverá ser capaz de resolver o problema da interoperabilidade semântica. Esses sistemas deverão ter habilidade para processar o modelo que o usuário faz do mundo e seus significados e processar também os modelos que há por trás das fontes de informação; para tal, o projeto de ontologias tem um papel de fundamental importância (Fonseca; Engenhofer; Borges, 2000). Ontologia é uma disciplina antiga – que vem desde Aristóteles – que estuda o ser e as suas propriedades. No domínio da Ciência da Informação e da Inteligência Artificial, uma das definições mais citadas diz que “ontologia é uma especificação formal e explícita de uma conceitualização compartilhada” (Grubber, 1996). A conceitualização é uma visão abstrata e simplificada de um domínio específico da realidade. Dessa forma, as ontologias são projetadas para possibilitar que o conhecimento seja compartilhado e reusado; elas explicam como um indivíduo, grupo, linguagem ou ciência entende um determinado domínio. Ontologia é um componente importante no processamento de conhecimento por computador e, conseqüentemente, para a solução de interoperabilidade semântica entre repositórios digitais federados, principalmente no mapeamento entre esquemas de representação distintos. A DAML+OIL e a Web Ontology Language (OWL), recentemente publicada pela W3C, são linguagens usadas para aplicações que necessitam compreender o conteúdo da informação, que são correntemente usadas na implementação de arquiteturas federadas (Martinez, 2006; Doerr, 2006). Experiências sobre o uso de ontologias para os problemas das bibliotecas digitais podem ser vistos em Weinstein (1998).

- *Agentes inteligentes* – o conceito de agente constitui uma poderosa ferramenta de abstração no

desenvolvimento de *softwares* que facilitem a construção de sistemas robustos, inteligentes e distribuídos. Um agente inteligente é um programa de computador especial capaz de ações flexíveis e autônomas num determinado ambiente, ou seja, ações reativas, proativas e sociais (ver em http://en.wikipedia.org/wiki/Software_agents). A solução de um problema complexo, como a federação de bibliotecas digitais, pode ser contemplada por um conjunto de agentes que interagem e cooperam entre si para alcançar os objetivos do sistema. Agentes inteligentes informacionais podem ser definidos como sistemas de *software* que acessam fontes de informação múltiplas, heterogêneas e geograficamente distribuídas no sentido de assistir ao usuário no processo de buscar informações relevantes (Martinez, 2006). As exigências mínimas que se colocam para os agentes informacionais são: saber onde encontrar os repositórios digitais; ter a capacidade de consultar esses repositórios de forma apropriada, interpretando, analisando e traduzindo a solicitação do usuário para cada fonte de informação; possuir métodos para processar os resultados da pesquisa, integrando os resultados em um formato comum e apresentando-os aos usuários (Birmingham, 1995; Tennant, 1998);

- *Digital Library Definition Language (DLDL)* – linguagem de especificação de bibliotecas digitais baseada em XML capaz de descrever API's para um grande número de repositórios digitais. A especificação é dividida em três seções: 1) as informações que as bibliotecas digitais contêm; 2) os métodos de acesso das bibliotecas digitais; 3) as informações a serem recuperadas das bibliotecas digitais (Zubair, 2000);

- *Encapsuladores, Tradutores, Adaptadores (Wrappers)* – são dispositivos de *software* que ocultam parte da heterogeneidade técnica e de modelo de dados de componentes nativos de sistemas de informação, tornando o acesso transparente para o sistema de mediação. De forma geral, os *wrappers* convertem dados provenientes de uma fonte de informação para um modelo de dados comum, e convertem consultas de aplicações em consultas específicas da fonte de informação correspondente. Por exemplo: um *wrapper* Z39.50 desempenha o papel de *gateway* entre o sistema mediador e um servidor Z39.50, com o qual se comunica via o protocolo nativo; nesse caso, o próprio Z39.50 (Melnik; Garcia-Molina; Paepcke, 2000; Bartelt et al., 2001; Baru et al., 1999).

- *CORBA*, sigla para *Common Object Request Broker* – especificação padrão criada pelo OMG

(Object Management Group), que propõe uma arquitetura de *software* para suportar a distribuição e garantir a interoperabilidade entre diferentes plataformas de *hardware* e sistemas operacionais, visando estabelecer e simplificar a troca de dados entre sistemas distribuídos e heterogêneos. A CORBA atua de forma que os objetos (componentes de *software*) possam comunicar-se de forma transparente em relação ao usuário, mesmo que seja necessário interoperar com outro *software* em outro diferente ambiente operacional. A especificação define um módulo intermediário entre o cliente e o objeto, o ORB (Object Request Broker). Trata-se de um *middleware* responsável em aceitar a requisição do cliente, localizar o objeto e passar os parâmetros necessários a esse objeto, fazer as chamadas dos métodos e entregar a resposta ao cliente. Dessa maneira, o usuário não precisa preocupar-se em saber onde esse objeto está localizado, em que sistema operacional ele roda ou qual programa foi usado para desenvolvê-lo. Esse enfoque de objetos distribuídos tem sido adotado como solução em algumas experiências importantes, como, por exemplo, no trabalho cooperativo entre CNRI e a Cornell University (Payette, 1999; Paepke et al., 1998).

- *Web services* (<http://www.w3.org/2002/ws/>) – é uma tecnologia utilizada na integração de sistemas e na comunicação entre aplicações diferentes. Essa comunicação efetiva-se de forma padronizada, possibilitando a independência de plataforma e de linguagem de programação. As bases para a construção de um *web service* são a linguagem XML e o protocolo SOAP (Simple Object Access Protocol) (<http://www.w3.org/TR/2003/REC-soap12-part0-20030624/>), definido pela W3C (World Wide Web Consortium). O transporte de dados é realizado via o protocolo HTTP (HyperText Transfer Protocol). Dessa forma, os dados são transferidos no formato XML e encapsulados pelo protocolo SOAP. Para ser utilizado por outras aplicações e para que tenha também seu funcionamento entendido pelos interessados, um *web service* deve ser publicado e disponibilizar uma definição de como ele é, como deve ser acessado e que valores retornará. Essas definições são descritas em um arquivo de acesso público em XML, de acordo com a padronização Web Service Description Language (WSDL) (Pamplona, 2004). *Web services* são blocos para construir aplicações, a partir dos quais se pode vislumbrar uma infra-estrutura para bibliotecas digitais distribuídas que disponibilize uma gama de funcionalidades, como, por

exemplo, buscas distribuídas, perfis de usuários, serviços de autenticação/autorização, diretório de coleções. Isso é possível porque a arquitetura de bibliotecas digitais é muito similar à arquitetura exigida para o estabelecimento de *web services*:

De fato, a infra-estrutura de *web services* pode ser pensada como uma biblioteca digital distribuída voltada para serviços ao invés de informação. Isto significa que muitas das questões em que a comunidade de bibliotecas digitais está envolvida, tais como metadados para descoberta de recursos, autenticação e autorização e modelos de negócio para o acesso a propriedade intelectual, são também aplicáveis a *web service* e devem ser resolvidas dentro de um contexto *web services*. (Gardner, 2001, p. 1).

- É importante notar que já existem pacotes de *software* para repositórios digitais que implementam interfaces como *web services*, como por exemplo o Fedora (<http://www.fedora.info/>); e o protocolo SRW (Search Retrieval Web Service) (<http://www.loc.gov/standards/sru/srw/index.html>), que tem como proposta ser aderente ao padrão de interoperabilidade estabelecido pelo Web Services Interoperability.

- *Arquitetura Peer-to-Peer (P2P)* – é uma arquitetura de sistemas distribuídos, na qual cada componente, chamado nó, é autônomo e tanto presta como fornece serviços aos outros nós. Essa arquitetura tornou-se muito comum com a popularização dos sistemas de troca de arquivos MP3, como era originalmente o Napster e são, atualmente, o Kazaa e o Emule. Aplicada a uma federação de bibliotecas digitais essa arquitetura, dada uma solicitação de serviço por determinado nó, uma busca, por exemplo, o nó ou nós capazes de melhor atender a busca são identificados segundo diferentes critérios e se encarrega de prestar o serviço. Bibliotecas digitais como BRICKS - Building Resources for Integrated Cultural Knowledge Services (<http://www.brickscommunity.org/>) funcionam segundo essa arquitetura.

- *Arquitetura multicamadas (multi-tier architecture)* - é uma nova dimensão para a arquitetura cliente-servidor, na qual uma aplicação é executada por mais de um diferente agente de *software*. Por exemplo, uma aplicação que usa *middleware* para um serviço de requisição de dados entre um usuário e uma base de dados emprega uma arquitetura multicamadas em três

módulos, ou seja, uma arquitetura de três camadas - camada de apresentação, camada lógica e camada de dados. Um número considerável de novos métodos e tecnologias provenientes da área de federação de bases de dados e sistemas de mediação estão sendo aplicados na integração de bibliotecas digitais. Essas experiências têm em comum uma arquitetura multicamadas que provê adaptadores e/ou wrappers para que fontes de dados distribuídas possam lidar com o problema de homogeneização de interface e representação de dados (Schallehn; Ending, 2000).

DIREITOS AUTORAIS

Copyright é a mais irritante barreira no desenvolvimento das bibliotecas digitais (Chepesuik, 1997, p.49).

Considerando que, de forma geral, uma federação de bibliotecas digitais oferece acesso a diversos repositórios digitais, um passo importante na consolidação de modelos de federação é a definição dos papéis dos usuários e dos seus direitos de acesso. Há, entretanto, um consenso absoluto por parte de toda a comunidade envolvida de que a gestão de direitos é um dos mais complexos e desafiadores problemas que a área de bibliotecas digitais tem que enfrentar. Discutir direitos conduz forçosamente para o território legal e o de negócios, os quais as bibliotecas, cuidadosamente, procuraram evitar no passado (Coyle, 2004a). As questões de direitos autorais (*copyright*) e propriedade intelectual foram sempre um problema difícil para as bibliotecas. Elas sempre tiveram que se equilibrar entre as leis de *copyright*, o conceito de *fair use*, ou seja, de uso razoável ou uso aceitável, e a doutrina da primeira alienação (*first sale doctrine*). O *fair use*, numa definição bem simples, permite que obras sejam copiadas para propósitos educacionais e empréstimo entre bibliotecas; por sua vez, a doutrina da primeira alienação tornou possível a existência das bibliotecas públicas de pesquisa modernas (i-DLR, 2003). Com o surgimento das bibliotecas digitais e do *e-commerce* e das novas configurações do mercado de conteúdo, essa questão torna-se crítica, pois se constata que o conceito tradicional de direito autoral não se ajustou no ambiente digital, dado que o controle de cópias, de integridade e acesso foi perdido: os objetos digitais são menos fixáveis, facilmente copiados e remotamente acessíveis por múltiplos usuários simultaneamente.

A maioria dos idealizadores de projetos importantes de bibliotecas digitais está consciente de que existem problemas de propriedade intelectual que devem ser resolvidos para que as bibliotecas digitais desenvolvam suas potencialidades plenamente. Algumas propostas expressam a intenção de resolver essas questões como parte do planejamento geral dos seus sistemas, embora sem muita especificidade de como isso possa ser praticado efetivamente. (Samuelson, 1995).

Os sistemas de bibliotecas e arquivos digitais atuais têm servido bem o seu público dentro dos limites da mídia impressa. É necessário, agora, estender esses modelos aonde for possível, e inventar novos modelos onde for necessário, para que se possa oferecer acesso aos artefatos digitais. De forma geral, as bibliotecas digitais necessitam de licenças flexíveis e inovadoras que lhes permitam, de forma legal, criar arquivos e coleções, gerar serviços compatíveis com as necessidades atuais e futuras de seus usuários e praticar estratégias apropriadas de preservação digital. O desafio atual é estabelecer os papéis, os direitos e as responsabilidades das bibliotecas e arquivos no que concerne à disponibilização do acesso público à informação digital. Pesquisas contínuas são necessárias para dar às bibliotecas a capacidade de gerenciar propriedade intelectual e proteger esses direitos sem inibir o legítimo acesso dos usuários aos materiais qualificados com esses direitos.

Embora seja comum para os utópicos do mundo tecnológico declarar a Internet, e por analogia todas as mídias digitais, uma zona livre de *copyright*, para o bem ou para o mal, as leis de direitos autorais aplicam-se a todas as formas de propriedade intelectual, estejamos ou não de acordo com elas (Coyle, 2004a). O problema para as bibliotecas é que, diferentemente de um negócio privado ou de uma editora, na maioria das vezes, elas são simplesmente custodiantes da informação – as bibliotecas não detêm os direitos sobre o material de seu acervo. É improvável que bibliotecas possam livremente digitalizar e prover acesso a materiais detentores de *copyright* da sua coleção. Ao invés disso, terão que desenvolver mecanismos para gerenciar esses direitos, procedimentos que permitam que elas disponibilizem informação sem violar as regras do direito autoral e da propriedade intelectual - tais procedimentos são chamados coletivamente de gestão de direitos

autorais. A indústria de conteúdos, especialmente a de entretenimento, vem desenvolvendo tecnologias que permitam um controle mais severo no ambiente digital. É necessário, entretanto, observar que as soluções para a mídia de entretenimento terão grande efeito sobre bibliotecas. Esses efeitos vão impactar as bibliotecas de duas formas: como consumidoras de produtos de informação comercialmente produzidos e como distribuidoras de recursos para o público em geral (Coyle, 2004b).

DRM é a sigla para *Digital Right Management* ou Gerenciamento de Direitos no Ambiente Digital (tradução dos autores), termo usado para as tecnologias que controlam como os conteúdos digitais são acessados e usados. Enquanto os detentores de direitos autorais têm o direito exclusivo sobre uma obra – tais como o direito de produzir uma cópia ou de distribuí-la para o público – na maioria das vezes eles não têm o direito de controlar como a obra será usada (por exemplo, o direito de ver uma obra ou de ler um trabalho). Os proprietários de conteúdos vislumbram as tecnologias de DRM como um meio de controlar o uso dos seus conteúdos; por outro lado, os usuários e os bibliotecários as avaliam como uma ameaça ao conceito de *fair use* (ALA, 2003).

No contexto de mudanças, é necessário observar o evidente interesse da academia por modelos abertos de publicação, como o Open Archive Initiative (OAI) e o Budapest Open Access Initiative (BOAI), como resposta ao custo crescente dos periódicos comerciais e práticas restritivas de publicação. Os modelos de publicações abertas certamente exigem novas estratégias de DRM que enfatizem o *fair use*, a proteção da propriedade intelectual de usos inadequados e os modelos de subscrição múltiplas, que incluam acessos taxados e não taxados. Além do mais, a comunidade envolvida na área de Pesquisa e Ensino (P&E) considera que as soluções comerciais de DRM afetam o frágil equilíbrio entre os que controlam os direitos autorais e os usuários, em favor dos primeiros, quando questionam os conceitos de *fair use* e *first sale* – dois princípios críticos e preciosos no contexto da P&E. Existe ainda a preocupação – bastante pertinente – de que algumas implementações de DRM comprometam a privacidade dos usuários (Martin et al., 2002).

Em face dessa realidade, há um esforço em andamento, no âmbito das comunidades de pesquisa em rede e bibliotecas digitais, para desenvolver soluções

inovadoras de DRM a fim de dar apoio às atividades de ensino e pesquisa e aos novos modelos de publicações acadêmicas. Martin e seus colaboradores relatam (Martin et al., 2002) um projeto que tem como foco específico a apresentação de uma arquitetura de referência voltada para a implementação de um sistema de DRM Federado (FDRM). O objetivo desse projeto é dar apoio ao compartilhamento de recursos em nível local e interinstitucional de uma forma discricionária, segura e com privacidade, enquanto se esforça para manter o equilíbrio entre os direitos do usuário e os dos proprietários dos conteúdos. A solução deve estar em consonância com as exigências das áreas de P&E, e devem incluir: acomodação dos aspectos intensamente colaborativos e distribuídos da maioria das atividades do mundo acadêmico; apoio ao *fair use* de materiais protegidos por *copyright* destinados a propósitos educacionais; suporte ao acesso diferenciado e granular aos recursos; prevenção ao mau uso dos recursos; garantia da integridade do recurso; e interoperabilidade com as infra-estruturas existentes e as emergentes.

Os requisitos preconizados pelos acadêmicos revelam uma distinção em relação ao modelo convencional de DRM – o modelo voltado para o comércio eletrônico, cuja função é primordialmente proteger os direitos do proprietário –, indicando que um conceito mais abrangente possa emergir incorporando gestão de acesso, bem como gestão dos direitos de propriedade intelectual e a preocupação com os direitos do usuário, assim como o direito dos proprietários. Dada essa diferença na interpretação, é possível verificar que o termo *DRM*, embora tenha sido adequadamente apropriado pela indústria de publicações, talvez não seja adequado para expressar os objetivos da comunidade acadêmica que poderiam ser mais bem traduzidos por outro termo (Martin, 2002). Novas formas de direito de uso também vêm sendo criadas e utilizadas em diversas instâncias e devem ser seriamente consideradas como alternativas aos modelos tradicionais, como a conhecida licença Creative Commons, (<http://creativecommons.org/>), e mais recentemente a Science Commons, (<http://sciencecommons.org/>). Creative Commons é um tipo de licença flexível de *copyright* para obras intelectuais surgida com a Internet. Ela abre a possibilidade de publicar e disponibilizar na rede os mais diferentes tipos de trabalho intelectual, de modo a permitir a cópia e reuso desses conteúdos por terceiros sob determinadas condições. Ao invés da tradicional enunciação de

copyright e todos os direitos reservados, a licença Creative Commons pretende permitir a cópia ampla, reuso, modificação, desenvolvimento e ampliação do trabalho intelectual original, desde que sejam garantidos alguns direitos. A partir disso, existem gradações, incluindo a redistribuição, alteração, uso comercial, entre outros. Há derivações da licença Creative Commons para trabalhos científicos, o Science Commons (<http://sciencecommons.org/>), e para recursos educacionais, o Open Educational Resources (<http://oercommons.org/>), que permitem acesso aberto aos recursos correspondentes (Korn; Oppenheim, 2006).

PARAMETROS DE AVALIAÇÃO

Um dos principais problemas em relação à interoperabilidade é que comparar soluções é uma tarefa muito difícil, posto que diferentes abordagens operam baseadas em concepções distintas e com objetivos algumas vezes conflitantes. Mesmo assim, é importante estabelecer alguns critérios para avaliação de soluções de interoperabilidade e compreender os *trade-offs* entre eles; no mínimo, isso nos ajuda a compreender melhor o problema de bibliotecas digitais distribuídas e interoperáveis. Paepcke et al., (1998) destacaram, entre tantos, seis critérios, cujas interpretações são as seguintes:

- a) Grau de autonomia – é um dos critérios mais críticos, pois tem forte impacto nos processos internos das organizações. Refere-se à quantificação da conformidade com as regras globais, que é exigida de cada componente participante do sistema, por exemplo, um serviço. Um alto grau de autonomia é desejável, pois pressupõe maior controle local sobre a implementação e a operação do componente e torna mais fácil a inclusão de sistemas legados como componente participante. Entretanto, a autonomia total pode gerar um descomprometimento com as regras globais – interfaces, protocolos de interação e formatos de dados não padronizados que podem ser arbitrariamente mudados. No outro extremo, os participantes do sistema têm que obrigatoriamente engajar-se nos procedimentos globais do sistema. Sistemas interoperáveis que funcionam na prática recaem sobre esses dois extremos.
- b) Custo da infra-estrutura – corresponde ao custo da infra-estrutura necessária para dar sustentação a uma solução, e para que um componente de serviço possa incorporar-se ao sistema. Corresponde ao custo de infra-estrutura necessário para dar sustentação ao sistema e para a incorporação de novos participantes.
- c) Facilidade de incorporação de componentes – esse critério, diferentemente do anterior, refere-se ao custo incremental de tornar interoperável um novo componente de serviço. Esse custo incremental pode envolver investimentos em hardware, ou pode estar expresso na forma de complexidade do software necessário para assegurar a interoperabilidade.
- d) Facilidade de uso – esse critério refere-se a dois itens: a complexidade de se criar um *software* cliente para um componente de serviço, ou seja, uma interface, e a complexidade de interagir com o componente quando em execução.
- e) Dimensão da complexidade das tarefas suportadas – está relacionado à capacidade do sistema de desenvolver, gerenciar, tornar interoperáveis e operar componentes que tenham como substrato tecnologias e padrões de alto nível de sofisticação.
- f) Escalabilidade – mede a capacidade do sistema de absorver crescentemente novos componentes de serviço.

Dependendo da dimensão que se deseja mensurar, outros itens podem ser incorporados na matriz de avaliação, incluindo parâmetros que avaliem impactos sociais, como nível de inclusão e impactos nos resultados de pesquisa; entretanto será sempre uma tarefa difícil de cumprir com precisão.

CONSIDERAÇÕES FINAIS

O estudo e o desenvolvimento de bibliotecas digitais, especialmente da sua vertente considerada por muitos especialistas como a mais sofisticada, que é a interoperabilidade, colocam em evidência, trazem para a prática e para o convívio com as tecnologias atuais

quase todos os fundamentos da Ciência da Informação. A pesquisa na área de bibliotecas digitais, pelo seu domínio amplo e pela sua complexidade, coloca-nos face a face com problemas que quase sempre só tivemos oportunidade de lidar no plano das idéias, e nos desafia a todo instante a recondicioná-los para uma nova realidade. Mas, tal qual a esfinge, as questões da área de bibliotecas digitais desafiam-nos com seu olhar de pedra, e são ainda muitos os enigmas que teremos que decifrar. Se não, vejamos alguns deles:

- Federação

A próxima geração de bibliotecas digitais deverá ser formada por sistemas de menor dimensão, de natureza diversificada (arquivos, museus, galerias, etc.), autônomos e independentes administrativamente, cada qual oferecendo – por vários meios e tecnologias - funcionalidades diferentes e acesso a diferentes conteúdos. A federação deve trabalhar com diferentes níveis de interoperabilidade, operando simultaneamente, como o caso de uso discutido no início deste artigo, que citava o problema de integrar conteúdos em cultura brasileira e língua portuguesa sobre “Brasil colônia e a influência da cultura negra”. A federação dessas bibliotecas digitais preencherá as necessidades de espaços informacionais semanticamente ricos e interconectados, dando margem ao surgimento de serviços de informação inovadores.

- Interface única e transparência para o usuário

Requer uma interface unificada que oculte toda a complexidade intrínseca à federação de bibliotecas digitais e proporcione um mecanismo universal para mapear e integrar cada repositório digital nessa interface, dando ao usuário a impressão de uma biblioteca única; a localização geográfica de cada fonte de informação deve ser indiferente para o usuário.

- Gestão de direitos

Para tornar viável o acesso universal aos estoques informacionais das bibliotecas

digitais acadêmicas, é necessário ampliar: os modelos atuais de gestão de direitos, quando for possível, e desenvolver novos modelos que preservem os conceitos de *fair use* e da *first sale doctrine*, essenciais para o desenvolvimento científico em escala global; desenvolver modelos automatizados de DRM que considerem, além dos direitos dos proprietários de materiais protegidos por *copyright*, os direitos de acesso individuais e institucionais dos usuários, preservando sua privacidade.

- Expansibilidade

As bibliotecas digitais deverão ser sistemas sempre em expansão; para tanto é necessário criar arquiteturas abertas e flexíveis que possibilitem a criação e a incorporação contínua de novos componentes de serviços. Todas as funcionalidades das bibliotecas deverão ser particionadas em um conjunto de serviços bem definidos, autodescritos, autoregistráveis e autoconfiguráveis, de forma que possam ser (semi-) automaticamente registrados e configurados ao serem “ligados” ao sistema na forma *Plug-and-Play* (Martinez, 2006). A integração das bibliotecas deve ser suficientemente flexível para permitir que cada biblioteca possa, individualmente, adicionar e/ou modificar características e componentes de serviço;

- Interoperabilidade semântica e sintática

Dada a natureza heterogênea da nova geração de bibliotecas digitais, a interoperabilidade deve ser a preocupação central dos requisitos dos sistemas. Os metadados e os *softwares* de interfaces devem ser automaticamente mapeados, de forma que a heterogeneidade sintática e a semântica possam ser resolvidas. As ontologias têm um papel-chave na resolução da interoperabilidade semântica e as pesquisas nessa área vão proporcionar a infra-estrutura necessária para o atendimento a essa necessidade, especialmente no que tange à heterogeneidade semântica (Martinez, 2006).

- Arquitetura e sistemas de mediação

É de fundamental importância para as pesquisas em interoperabilidade de repositórios digitais explorar concepções novas de arquiteturas para sistemas de mediação para federação, bem como tecnologias da área de T.I. Agentes inteligentes, arquitetura multicamadas (*multi-tier architecture*), CORBA, web-services, wrappers são tecnologias potencialmente importantes para o desenvolvimento de *middleware* para federação de bibliotecas digitais.

- Aspectos sociais, políticos e culturais

Por último, mas não menos importante, estão as pesquisas que investigam os impactos da federação de acervos digitais na educação e na cultura, mas precisamente como ferramenta básica nas áreas de educação e treinamento a distância; no apoio aos programas de inserção digital e de cidadania; e como meio para potencializar a visibilidade digital de manifestações culturais e artísticas.

REFERÊNCIAS

- ALA-American Library Association. *Digital right management and libraries*. Washington, DC: ALA, 2003. Disponível em: <<http://www.ala.org/ala/washoff/WOissues/copyrightb/digitalrights/digitalrightsmanagement.htm>>. Acesso em: 15 mar. 2007.
- ARMS, W.Y. Thoughts about interoperability in the NSDL: draft for discussions. August 2000. Disponível em: <<http://www.cs.cornell.edu/wya/papers/NSDL-Interop.doc>>. Acesso em: 15 mar. 2007.
- ARMS, W.Y. et al. A spectrum of interoperability: the site for science for prototype for the NSDL. *D-Lib Magazine*, v.8, n.1, 2002. Disponível em: <<http://www.dlib.org/dlib/january02/arms/01arms.html>>. Acesso em: 4 jan. 2007.
- BARTELT, A. et al. Buiding infrastructure for digital library. In: DELOS NETWORK OF EXCELLENCE WORKSHOP ON INTEROPERABILITY AND MEDIATION IN HETEROGENEOUS DIGITAL LIBRARIES, 3., 2001, Darmstadt.. *Proceedings*. European Research Consortium For Informatics And Mathematics, 2001. 5 p. Disponível em: <<http://www.ercim.org/publication/ws-proceedings/DelNoe03/6.pdf>>. Acesso em: 18 mar. 2007.
- BARU, C. et al. XML-based information mediation for digital libraries. In: ACM CONFERENCE ON DIGITAL LIBRARIES, 4., 1999, Berkeley. *Proceedings...* ACM, 1999. p.214-215. Disponível em: <<http://www.npaci.edu/DICE/Pubs/dl99-demo.pdf>>. Acesso em: 18 mar. 2007.
- BERNERS-LEE, T., HENDLER, J., LASSILA, O. The semantic web. *Scientific American*, May, 2001. Disponível em: <<http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>>. Acesso em: 24 abr. 2001.
- BIRMINGHAM, W. An agent-based architecture for digital libraries. *D-Lib Magazine*, July 1995. Disponível em: <<http://www.dlib.org/dlib/July95/07birmingham.html>>. Acesso em: 7 mar. 2006.
- CHEPESUIK, R. The future is here: America's Libraries go Digital. *American Libraries*, v.2, n.1, p.47-49, 1997.
- COYLE, K. The "Rights" in the digital rights management. *D-Lib Magazine*, v.10, n.9, 2004a. Disponível em: <<http://www.dlib.org/dlib/september04/coyle/09coyle.html>>. Acesso em: 15 mar. 2007.
- COYLE, K. Rights Management and digital library requirements. *Ariadne*, n.40, 2004b. Disponível em <<http://www.ariadne.ac.uk/issue40/coyle>>. Acesso em: 15 mar. 2007.
- DOERR, M. Increasing the Power of Semantic Interoperability for the European Library. *Ercim News Online Edition*, 2006. Disponível em: <http://www.ercim.org/publication/Ercim_News/enw66/doerr.html>. Acesso em: 6 mar. 2007.
- FONSECA, F.; ENGENHOFER, M.; BORGES, K. Ontologias e interoperabilidade semântica entre SIG's. In: GEOINFO 2000 - WORKSHOP BRASILEIRO DE GEOINFORMÁTICA, 2., 2000, São Paulo. *Anais...* Disponível em: <<http://www.tecgraf.puc-rio.br/geoinfo2000/anais/011.pdf>>. Acesso em: 8 mar. 2007.
- GARDNER, T. An introduction to web services. *Ariadne*, v.29, 2001. Disponível em: <<http://www.ariadne.ac.uk/issue29/gardner/>>. Acesso em: 8 mar. 2007.
- GRUBER, T. *What is an ontology: knowledge system al laboratory*, 1996. Disponível em: <<http://www.ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 18 mar. 2007.
- HATALA, M. et al. The interoperability of learning object repositories and services: standards, implementations and lessons learned. In: INTERNATIONAL WORLD WIDE WEB CONFERENCE, 13., 2004, New York. *Proceedings of the Conference on Alternate track papers & posters*. New York, ACM, 2004. p. 19 - 27. Disponível em: <<http://www.sfu.ca/~mhatala/pubs/www04-interop-final.pdf>>. Acesso em: 19 mar. 2007.
- i-DLR – Interactive Digital Library Resource Information System. *Copyright issues and intellectual properties rights in digital libraries*. 2003. Disponível em: <<http://www.coe.missouri.edu/~DL/iDLR/viewpaper.php?pid=20>>. Acesso em: 15 mar. 2007.

- KORN, N.; OPPENHEIN, C. Creative Commons Licences in Higher and Further Education. *Ariadne*, n.46, 2006. Disponível em: <<http://www.ariadne.ac.uk/issue49/korn-oppenhein>>. Acesso em: 11 jul. 2007.
- MARCONDES, C.H.; SAYÃO, L.F. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. *Ciência da Informação*, v.30, n.3, p.24-33, 2001. Disponível em: <<http://www.scielo.br/pdf/ci/v30n3/7283.pdf>>. Acesso em: 19 mar. 2007.
- MARTIN, M. et al. Federated Digital Rights Management. *D-Lib Magazine*, v.8, n.7/8, 2002. Disponível em <<http://www.dlib.org/dlib/july02/martin/07martin.html>>. Acesso em 15 mar. 2007.
- MARTINEZ, R.J.F. Agents and ontologies working together to federate digital libraries. *TCDL Bulletin*, v.2, n.2, 2006. Disponível em: <<http://www.ieee-tcdl.org/Bulletin/v2n2/rodriguez-martinez/rodriguez-martinez.html>>. Acesso em: 15 fev. 2007.
- MELNIK, S.; GARCIA-MOLINA, H.; PAEPCKE, A. A Mediation infrastructure for digital library services. In: ACM CONFERENCE ON DIGITAL LIBRARIES, 5., 2000, San Antonio. *Proceedings...* ACM, 2000. p.123-132. Disponível em: <<http://infolab.stanford.edu/~melnik/pub/dl00.pdf>>. Acesso em: 18 mar. 2007.
- MILLER, P. Interoperability. What is it and why should I want it? *Ariadne*, n.24, 2000. Disponível em: <<http://www.ariadne.ac.uk/issue24/interoperability/>>. Acesso em: 18 mar. 2007.
- NISO - NATIONAL INFORMATION STANDARDS ORGANIZATION. *Z39.50 – A primer on the protocol*. Bethesda, MD: NISO Press, 2002. Disponível em: <http://www.niso.org/standards/resources/Z3950_primer.pdf>. Acesso em: 17 abr. 2007.
- ONLINE DICTIONARY FOR LIBRARY AND INFORMATION SCIENCE - (ODLIS). 2004. Disponível em: <<http://lu.com/odlis/>>. Acesso em: 7 jul. 2008.
- PAEPCKE, A. et al. Interoperability for digital libraries worldwide. *Communication of the ACM*, v.41, n.4, p.33-43, 1998.
- PAMPLONA, V.F. *Web Services: construindo, disponibilizando e acessando web services via J2SE e J2ME*. Javafree.org, 2004. Disponível em: <<http://www.javafree.org/contentview.iff?idContent=4>>. Acesso em: 08 mar. 2007.
- PAYETTE, S. et al. Interoperability for digital objects and repositories: the Cornell/CNRI Experiments. *D-Lib Magazine*, v.5, n.5, 1999. Disponível em: <<http://dlib.org/dlib/may99/payette/05payette.html>>. Acesso em: 7 fev. 2007.
- PIRRI, M.; PETTENATI, M.C.; GIULI, D. Design of a federation service for digital libraries: the case of historical archives in the PORTA EUROPA Portal (PEP) Pilot Projec. In: DC-2002: METADATA FOR E-COMMUNITIES, Florence. *Proc. Int. Conf. on Dublin Core and Metadata for e-Communities 2002*. Florence: Firenze University Press, 2002. p.157-162. Disponível em: <<http://www.bncf.net/dc2002/program/ft/paper18.pdf>>. Acesso em: 18 mar. 2007.
- SAMUELSON, P. Copyright and digital libraries. *Communication of the ACM*, v.38, n.3, April 1995. Disponível em: <<http://doi.acm.org/10.1145/205323.205324>>. Acesso em: 15 mar. 2007.
- SAYÃO, L.F.; MARCONDES, C.H. *Guia de software em automação de bibliotecas*. Brasília: MEC/SEu/PNU, 1989.
- SCHALLEHN, E.; ENDING, M. *Using source capability description for the integration of digital libraries*. 2000. Disponível em: <<http://www.is.informatik.uni-kiel.de/~hjk/wsgdb00/Paper/Schallengdig.os>>. Acesso em: 8 mar. 2007.
- SRW: Search/Retrieve Web Service. Disponível em: <<http://www.loc.gov/standards/srw/index.html>>. Acesso em: 25 mar. 2007.
- SURE, Y.; STUDER, R. *Semantic web technology for digital libraries*, 2005. Disponível em: <http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2005_sw_for_dl.pdf>. Acesso em: 2 maio 2007.
- TENNANT, R. *Interoperability: the holy grail*. *Library Journal*, 1998. Disponível em: <<http://www.libraryjournal.com/article/CA156495.html>>. Acesso em: 5 jan. 2007.
- THE OPEN ARCHIVES INITIATIVE PROTOCOL FOR METADATA HARVESTING. Protocol Version 2.0 of 2002-06-14. Disponível em: <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>. Acesso em: 22 mar. 2007.
- UKOLN. *Interoperability focus: looking at interoperability*. 2005. Disponível em: <<http://www.ukoln.ac.uk/interop-focus/about/leaflet.html>>. Acesso em: 18 mar. 2007.
- WARREN, P.; THURLOW, I.; ALSMEYER, D. Applying semantic technology to a digital library. In: DAVIS, J.; STUDER, R.; WARREN, P. *Semantic Web technologies: trends and research in ontology-based systems*. Sussex: John Wiley, 2006. p.237-257.
- WEINSTEIN, P.C. Ontology-based metadata: transforming the MARC legacy. In: ACM conference on Digital Libraries 3., Pittsburgh, *Proceedings...* Pittsburgh: ACM, 1998. p.254-263.
- ZUBAIR, M. et al. Dynamic construction of federated digital libraries. In: INTERNATIONAL WORLD WEB CONFERENCE THE WEB: THE NEXT GENERATION, 9., 2000, Amsterdam. *Proceedings...* Disponível em: <<http://www9.org/final-posters/poster17.html>>. Acesso em: 2 mar. 2007.