

Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes - URN, PURL, DOI, Handle System, CrossRef e OpenURL

Digital libraries' interoperability: The role of persistent identifiers' systems - URN, PURL, DOI, Handle System, CrossRef and OpenURL

Luís Fernando SAYÃO¹

RESUMO

O conceito de repositórios digitais distribuídos e interoperáveis - sejam eles bibliotecas ou arquivos digitais - para ser viável, depende fortemente de uma infra-estrutura de identificação persistente e da confiabilidade do esquema de *links* entre recursos informacionais distintos, porém relacionados, de forma a assegurar ao usuário o acesso contínuo e conveniente aos conteúdos digitais. Neste trabalho, descreve-se o atual estado da arte da identificação persistente na Internet e da vinculação por *links*, tendo como foco a interoperabilidade das bibliotecas digitais de arquitetura aberta. Contrastam-se as diversas abordagens e analisam-se os mais importantes esquemas e o relacionamento entre eles: URN, PURL, DOI, CrossRef, Handle System and OpenURL.

Palavras-chave: bibliotecas digitais; identificadores persistentes; URI; URN; PURL; DOI; CrossRef; Handle System; OpenURL.

ABSTRACT

To be feasible, the concept of distributed and interoperable digital repositories - libraries or archives - depends strongly on an infrastructure of persistent identification, as well as on the reliability of the linkage scheme between distinct but related information resources, in order to assure to the user a continuous and convenient access to the digital contents. This paper describes the current state-of-the-art of persistent identification and linkage in the Internet, focusing on the interoperability of the open digital libraries. It contrasts various approaches to the issue and analyzes the most important schemes and their relationship: URN, PURL, DOI, CrossRef, Handle System and OpenURL.

Key words: digital libraries; persistent identifiers; URI; URN; PURL; DOI; CrossRef; Handle System; OpenURL.

INTRODUÇÃO

O conjunto de tecnologias, padrões e protocolos que viabiliza e sustenta a Web tal como a conhecemos

como pessoas comuns, vem também gerando exponencialmente condições técnicas que possibilitam uma revolução contínua, que se reconfigura a cada momento, em termos de disponibilidade e acessibilidade

¹ Doutor em Ciência da Informação. Consultor Técnico, Comissão Nacional de Energia Nuclear, Centro de Informações Nucleares. Rua General Severiano, 90, Botafogo, 22294-900, Rio de Janeiro, RJ, Brasil. E-mail: <lsayao@cnen.gov.br>. Recebido em 28/9/2006 e aceito para publicação em 9/2/2007.

de recursos informacionais em formato digital. Esta oferta de recursos digitais, movida cada vez mais pela força do mercado e pelas necessidades da sociedade do conhecimento, é dirigida a um universo cada vez mais amplo e diversificado de usuários, em escala quase planetária. O encontro de tecnologias e padrões abertos, além de transformar radicalmente a forma pela qual os disseminadores tradicionais de informações geram e distribuem produtos e serviços de informação para seus clientes, e de mudar profundamente a forma como as organizações de conhecimento interagem mutuamente, guarda ainda em si um potencial de transformações cujas perspectivas são inimagináveis.

Apoiados pela tecnologia corrente, os recursos informacionais das bibliotecas digitais, por exemplo, podem ser disseminados através de redes abertas; enquanto na outra ponta, indiferente à tecnologia que gerou esses recursos, o usuário é capaz de recuperar e manipular essas informações sem a necessidade de *softwares* clientes especiais ou de sistemas proprietários, componentes essenciais por muito tempo. O usuário necessita apenas, como instrumento de busca e recuperação, dos navegadores *web* (*web browsers*) comuns - *softwares* clientes, genéricos, familiares e de múltiplos propósitos, já instalados nos computadores dos usuários para as mais diversas finalidades e cujo traço mais significativo é filtrar na tela do usuário a complexidade crescente da Web.

Outro aspecto importante, viabilizado pela infraestrutura tecnológica e pela utilização de padrões abertos, é a interoperabilidade dos repositórios digitais, cuja viabilidade técnica dá margem à realização, de fato, do conceito de repositórios digitais distribuídos - vasto conjunto de estoques de recurso digitais sediados e mantidos por organizações distintas, que, no entanto, se unificam na perspectiva do usuário. Eles se traduzem efetivamente pela possibilidade de acessar recursos digitais, independente da sua localização geográfica, diretamente através do acionamento de *links* presentes em bases de dados, catálogos, índices e portais, assim como de *links* presentes em outros objetos digitais, como por exemplo, um artigo incluído numa publicação eletrônica.

O sucesso de um sistema de informações distribuídas - tal como é caracterizada a própria *World Wide Web* - depende fortemente da vinculação consistente entre os recursos que estão disponibilizados *on-line*. Isto se traduz pela estabilidade dos *links*

presentes nos catálogos, nos índices e nas listas que constituem os diversos serviços de descoberta de recursos. Contudo, para isso se efetivar, é necessário que para cada recurso seja atribuído um nome que os identifique permanentemente e sem ambigüidades e independente de localização.

A vinculação entre recursos pode variar de uma simples inserção de um *link* com o sentido de enriquecer uma informação - por exemplo, um *hiperlink* no nome de uma organização que é mencionada num documento -, até a citação formal de um outro recurso, expresso por uma referência bibliográfica. No domínio das bibliotecas digitais, a categoria de *links* que representa uma citação formal, isto é, a que formaliza uma vinculação de um recurso a outro, é a mais significativa para a sustentação da integridade do conteúdo intelectual, constituindo o que a literatura chama de "*link* referencial" (Caplan; Arms, 1999). É especialmente importante que estes *links* continuem a identificar e proporcionar acesso por longo prazo aos recursos que eles referenciam. Sua presença torna-se um fator importante e um desafio para as estratégias de arquivamento e preservação adotadas para os recursos cujos *links* tenham papel relevante (National Library of Australia, 2002).

Parece claro, portanto, que os *links* entre os recursos digitais cumprem um papel importante no estabelecimento da funcionalidade entre bibliotecas e arquivos digitais configurados de forma distribuída. Isto porque a interoperabilidade entre esses repositórios é garantida - entre outras coisas - pela confiança de que os enlaces entre os recursos por eles disponibilizados permaneçam estáveis por longo período de tempo. A condição de permanência dos *links* por longo prazo é um fator importante para a integridade dos objetos digitais, dado que estes recursos são compostos também pelos seus relacionamentos com outros objetos digitais; relacionamentos que, por sua vez, se realizam através de *links* acionáveis. Quanto à identificação permanente, ela garante o acesso contínuo ao próprio recurso, assegurando que este seja localizado e se "*materialize*" todas as vezes que seu *link* referencial seja acionado ou clicado; a identificação permanente constitui também um aspecto importante nas estratégias de preservação adotadas para os recursos digitais (National Library of Australia, 2002; Sayão, 2006).

Neste cenário, para o desenvolvimento pleno de bibliotecas e arquivos digitais de arquitetura aberta

e plenamente interoperáveis, há um interesse contínuo por parte dos vários atores envolvidos - que incluem principalmente as organizações responsáveis pelo ordenamento da Internet, as organizações internacionais de normalização e as organizações voltadas para o desenvolvimento de arquivos e bibliotecas digitais - de superarem a fragilidade dos esquemas atuais de identificação de recursos na web, através do estabelecimento de mecanismos persistentes de identificação que sejam mais robustos e independentes de fatores voláteis, como os vinculados ao endereço do recurso na Web e/ou nas tecnologias e protocolos de uso corrente (Payette, 1998). Tanto este problema, como as soluções propostas e adotadas, serão discutidos brevemente neste trabalho.

O Problema da identificação de recursos na Internet: nome versus localização

“Nomes são blocos de estrutura vitais para as bibliotecas digitais. Eles são necessários para identificar objetos digitais, para registrar propriedade intelectual vinculada a esses objetos e para registrar mudanças na propriedade dos objetos digitais. Eles são necessários para citação, para recuperação de informação, e são usados como links entre objetos” (Arms; Reston, 1995).

Para desempenhar tais papéis, os nomes têm que ser únicos e persistentes. Um nome persistente, no contexto dos repositórios digitais, é compreendido como um identificador único, o qual deverá perdurar por um período tão longo quanto seja necessário; mesmo que a organização que o atribuiu ao objeto não mais exista quando este for usado. Para este esquema funcionar, necessita-se estabelecer uma infra-estrutura administrativa cuja função será decidir quem pode assinalar nomes que identifiquem univocamente os recursos digitais de forma persistente. É imprescindível também criar sistemas automatizados capacitados a resolver nomes rapidamente, ou seja, sistemas que dêem como resposta a localização de onde está armazenado um objeto digital detentor de um dado nome (Arms; Reston, 1995).

Consideradas as funcionalidades que as estruturas de nomes persistentes e globalmente únicos

viabilizam - nas citações cruzadas, na recuperação de informação, no estabelecimento de links entre objetos e no gerenciamento de copyright -, conclui-se que tais estruturas são tão importantes para bibliotecas digitais quanto são os códigos International Standard Book Number (ISBN) para as bibliotecas tradicionais. Deve-se ressaltar, entretanto, que o ISBN se refere somente a um conjunto específico de recursos informacionais constituído pelos livros, enquanto as estruturas de nomes persistentes e unívocos devem identificar todos os recursos, incluindo os que aparecerão no futuro.

Nos dias de hoje, o URL - *Uniform Resource Locator* - é a porta de entrada para os conteúdos que estão disponíveis na Web. Ele define, como o seu próprio nome diz, a localização do recurso. Diferentemente do ISBN, número de chamada de um livro, que é raramente reclassificado, o maior problema com os URLs é que eles mudam com muita frequência: esta é precisamente a maior fragilidade desse esquema. A crítica mais contundente que se faz ao URL, é que ele mistura, numa mesma cadeia de caracteres, elementos que desejavelmente deveriam estar separados, tais como: o método pelo qual o recurso é acessado; o nome da máquina servidora e o caminho (*path*) do documento -isto é, sua localização e o nome do arquivo- que nem sempre é único (Cleveland, 1998). Apesar dessa fragilidade, freqüentemente tratamos o URL como se ele fosse um identificador formal que, de fato, representasse um objeto digital. Na realidade, o URL é simplesmente um endereço, usualmente mascarado como um identificador. Confiar nele como identificador único para os recursos digitais, é como usar o endereço residencial de uma pessoa para identificá-la, em vez de usar o seu CPF (Payette, 1998).

O efeito da instabilidade do URL está continuamente presente, tanto para quem busca informação profissionalmente, como para quem simplesmente navega na Web. Repetidamente estamos diante de *links* quebrados, e a cada dia cresce a proporção de links que, ao serem acionados, produzem como resposta a mensagem recorrente: “Erro HTTP - 404 Arquivo não encontrado”. Estima-se que esse erro ocorre em cerca de 7% dos *links* (Bigwood, 1999). Numa medida ao longo do tempo, tendo como objeto de estudo uma amostra aleatoriamente selecionada de URL's, foi demonstrado que somente cerca de 34% dos URL's permaneciam ativos depois de um período de quatro anos (Koehler, 2002).

Portanto, o funcionamento adequado e interoperável das bibliotecas digitais depende de um esquema global de identificadores únicos, que persistam além das instituições geradoras, e que não sejam atrelados a localizações e processos específicos. Os nomes devem permanecer estáveis, mesmo que os recursos correspondentes sejam transferidos de localização, de propriedade e/ou sejam migrados de mídia de armazenamento. A mobilidade, quase sempre imprevisível, dos recursos informacionais da Internet, é, na melhor das hipóteses, uma frustração para quem estava certo de ter encontrado o que procurava. Entretanto, para os profissionais de informação, essa mobilidade é um problema grave, o qual compromete os serviços oferecidos pelos repositórios digitais e impõe o ônus inaceitável da manutenção intensiva de seus catálogos e bases de dados. É necessário, portanto, o desenvolvimento de uma solução geral para este problema.

Porque os links param de funcionar

Os *links* param de funcionar se, por algum motivo, os recursos para os quais eles apontam são movidos, renomeados ou deletados. Os *links* “quebram”, quando os nomes de domínios, de arquivos ou diretórios que compõem os endereços dos recursos, sofrem algum grau de modificação. As razões mais comuns para que isso aconteça são as seguintes (National Library of Austrália, 2002):

- a) O recurso pode ter sido transferido, como resultado de reorganização no sistema de arquivos onde ele estava localizado, de mudanças na estrutura organizacional, ou de troca do nome de domínio; ou ainda, pode ter sido atribuído um novo URL ao recurso, como desdobramento da transferência de propriedade do recurso;
- b) O recurso pode ter sido removido porque está desatualizado ou porque perdeu o valor, seja comercial, intelectual, histórico etc., que lhe era atribuído ou percebido por seu proprietário. Neste caso o recurso, enquanto objeto digital, “desaparece”.

Embora um dado recurso tenha sido transferido ou perdido seu valor no âmbito da organização proprietária, na perspectiva do usuário o mesmo recurso pode continuar a ter valor permanente como referência científica, técnica, administrativa, legal ou histórica.

Além do mais, o URL desse recurso, por ter sido referenciado por diversos outros recursos, e também por catálogos, portais, diretórios, bibliografias, etc., ou ainda, por ter sido registrado no *bookmark* de instituições e indivíduos, continua a estabelecer falsas conexões.

Os sistemas de identificação persistente

Um identificador persistente é um nome para um recurso digital que permanece o mesmo para sempre, independente da localização do recurso. O uso de um identificador persistente assegura que, mesmo quando um documento é movido, ou sua propriedade é transferida, os *links* para ele permaneçam efetivamente acionáveis. Existe um espectro restrito de estratégias de implementação de identificadores persistentes, cuja adoção vai depender das circunstâncias técnicas, administrativas e políticas de cada instituição, sobretudo, da sua visão de futuro e do grau de interoperabilidade desejado com outros sistemas. As estratégias são as seguintes:

- a) Redirecionamento - é uma estratégia mínima, posto que utiliza os recursos padronizados do servidor *web* para redirecionar as solicitações para a posição corrente do recurso. Este método é difícil de se gerenciar quando se trata de *websites* de grande porte;
- b) Instalação de um resolvidor apoiado em banco de dados - pressupõe um software servidor de *links*, rodando sobre um banco de dados e tendo como finalidade, mapear a localização corrente do recurso, ou seja, o URL corrente. Uma opção nesta categoria é o software servidor PURL - *Persistent URL* - disponibilizado pela *Online Computer Library Center* (OCLC) (<http://purl.oclc.org/>) - que será objeto de análise na seção PURL-Persistent URL;
- c) Contratação de sistema de identificação persistente, oferecido por outra organização - existem vários sistemas de identificação persistente projetados para uso na Internet, baseados em padrões abertos, com objetivos e enfoques distintos. Por exemplo: *Digital Object Identifiers* (DOI), *Handle System* e também PURL, posto que a OCLC oferece serviço de identificação on-line para terceiros. Estes sistemas serão objetos de análise nas seções seguintes.

Mas não importa qual seja o sistema de identificação permanente considerado, ele só será efetivo

se for mantido. Quando recursos são movidos, a localização atual deve ser obrigatoriamente associada com o identificador permanente, seja qual for o sistema. Não existe sistema que não requeira tal gerenciamento.

Uniform Resource Name - URN

O problema de se criar uma estrutura que comporte identificadores únicos para recursos digitais é uma questão recursiva no domínio da comunidade Internet. A extensa documentação sobre a instabilidade dos links é uma expressão de o quanto isso preocupa a todos os envolvidos. A *Internet Engineering Task Force* (IETF) (<http://www.ietf.org>) - comunidade internacional aberta, composta por várias categorias de profissionais especialistas em Internet, cuja preocupação é a evolução e operação da Rede, incluindo os aspectos referentes aos padrões e protocolos, reconhecendo há muito tempo as limitações do URL instalou grupos de trabalhos que estabeleceram requisitos e uma estrutura para o desenvolvimento de esquema de nomes (Berners-Lee, 1988; Sollins, 1988; Petke; King, 1999).

O esquema de identificadores usados atualmente na Internet é a *Uniform Resource Identifiers* (URI), cuja especificação está descrita no documento publicado em 1988 "RFC 2396 - *Uniform Resource Identifiers (URI): Generic Syntax*" (Berners-Lee, 1988). Nesse documento se distinguem duas grandes categorias de URI's: *Uniform Resource Locator* (URL) e a *Uniform Resource Name* (URN). A RFC2396 estabelece que "um URI pode ainda ser classificado como um localizador, um nome ou ambos", sendo que o URL "se refere ao subconjunto do URI que identifica o recurso através da representação do seu mecanismo primário de acesso (por exemplo, sua localização na rede)". Enquanto o URN "se refere ao subconjunto do URI que é preciso permanecer globalmente único e persistente mesmo quando o recurso deixa de existir ou se torna indisponível" (Berners-Lee, 1988). Dessas duas categorias, a única que está plenamente desenvolvida e disseminada é o URL (Dack, 2001).

Um documento de especificação que, primordialmente, delineava os requisitos mínimos da URN, RFC 1737- *Functional Requirements for Uniform Resource Names*, foi lançado em dezembro de 1994 (Sollins, 1994). Este documento fixava a concordância geral em relação aos requisitos básicos para URN's e,

por definição, para qualquer outro esquema útil de identificadores persistentes. Esses requisitos são os seguintes (Sollins, 1994; Dack, 2001):

- a) *Escopo global*: o URN é um nome com validade em escala global, o que não implica em dependência relativa à localização. Ele tem o mesmo significado em qualquer lugar;
- b) *Unicidade global*: o mesmo URN não deverá nunca ser assinalado para dois recursos diferentes;
- c) *Persistência*: o tempo de existência de um URN deve ser permanente. Isto é, o URN será globalmente o mesmo para sempre, e será usado como referência a um recurso digital, para além da existência do recurso e da organização que lhe atribuiu o nome;
- d) *Escalabilidade*: o URN pode ser assinalado para qualquer recurso que venha a ser concebido para disponibilização na rede por centenas de anos;
- e) *Suporte aos sistemas legados*: o esquema deve permitir a incorporação dos sistemas de identificação já existentes - sistemas legados - que satisfaçam as demais exigências aqui descritas. Por exemplo, códigos ISBN e ISSN (International Standard Serial Number), identificadores ISO (International Organization for Standardization) e outros de igual importância;
- f) *Extensibilidade*: qualquer esquema de URN's deve permitir extensões futuras para o esquema;
- g) *Independência*: as condições sob as quais um nome pode ser atribuído são de responsabilidade única e exclusiva da autoridade que atribui nomes;
- h) *Resolução*: a URN não impede a resolução (tradução para um URL). Mais especificamente, para URN's que têm URL's correspondentes, deve existir algum mecanismo viável para traduzir um URN em um URL.

Sintaxe do URN

Após muita discussão no âmbito da IETF, foi lançado em maio de 1997 o RFC 2141-URN Syntax (Moats, 1997), que expressava o consenso em torno da sintaxe do URN. Essa sintaxe reconhecia a exigência de acomodar as necessidades da diversidade de usuários Internet e de diferentes esquemas existentes, tinha, portanto, que ser uma estrutura bastante simples: ela é composta de dois elementos:

urn:<Namespace Identifier (NID)>:<Namespace specific string (NSS)>

O primeiro elemento, o NID, identifica univocamente a autoridade designada para atribuir nomes - autoridade nomeadora -, que por sua vez pode autorizar sub-autoridades. A unicidade é assegurada pela exigência de que cada autoridade garanta unicidade no âmbito do seu conjunto de nomes - ou seja, no seu *namespace* - e que o NID seja aprovado e registrado pelo IETF. A cadeia de caracteres específica que forma o segundo elemento pode tomar qualquer forma determinada pela autoridade nomeadora, observando-se a unicidade no âmbito do seu *namespace* e as restrições em relação a alguns poucos caracteres especificados no documento RFC 2141 (Moats 1997)

Sistema de resolução

Há um consenso geral sobre uma distinção importante a ser feita quando se fala em esquemas padronizados de identificação. Esta distinção deve acontecer entre o “esquema de nomes” e os “sistemas de resolução”. Esquema de nomes: são políticas e procedimentos para a criação e atribuição de URNs dentro de um domínio particular. Teoricamente, qualquer pessoa, grupo ou organização, pode tornar-se uma autoridade nomeadora, a qual tem a responsabilidade única e exclusiva de implementar e gerenciar o esquema de nomes dentro do seu domínio. Por outro lado, o sistema de resolução é um sistema que traduz um URN particular em seu localizador, por exemplo, o URL. Uma vez que o identificador único tenha sido estabelecido para um objeto digital, ele deve ser registrado e armazenado num serviço acessível via rede que pode resolvê-lo em sua localização. Os esquemas de nomes devem ser independentes dos serviços de resolução: o URN não deve estar vinculado a nenhum sistema de resolução específico; entretanto, um serviço de resolução deve ser capaz de resolver um URN proveniente de qualquer esquema de nomes (Payette, 1998). Esta regra é determinante para a interoperabilidade entre sistemas de identificadores.

Perspectiva para o URN

Para o pleno funcionamento do esquema URN seria necessário, entre outras coisas, que fosse incorporada aos navegadores web padrões - Internet

Explorer, Netscape, etc. - a capacidade de reconhecer e direcionar os URNs. A IETF continua trabalhar em conjunto com toda a comunidade Internet, incluídos os desenvolvedores de navegadores web, no intuito de alcançar consenso sobre os padrões necessários para tal. Enquanto isso não acontece, todos os esquemas de URN já implementados, assim como os que o serão a curto e médio prazo, terão que utilizar uma intermediação, ou seja, terão que lançar mão de servidores *proxy*, ou seja, *alternativos*, que direcionam a requisição para o computador servidor onde está localizado o recurso que se deseja. Isto significa que, para funcionarem, os identificadores têm que estar encapsulados em URLs que os redirecionem ao endereço do servidor *proxy*. Veremos isto em alguns exemplos de implementações de identificadores persistentes, analisados a seguir: Persistent URL (PURL), Handle System e Digital Object Identifier (DOI).

PURL - Persistent URL

O PURL (<http://purl.oclc.org/>) é um esquema desenvolvido pela *Online Computer Library Center* (OCLC) (<http://www.oclc.org/>) - na tentativa de separar o nome de um recurso Internet, de sua localização, aumentando, dessa forma, a probabilidade de que o recurso esteja disponível quando o seu link for acionado. Nesse empreendimento, a OCLC tinha como perspectiva apoiar também o desenvolvimento e a aceitação da tecnologia URN (Shafer et al., 1996). O sistema de URL's Persistentes inicialmente visava reduzir os custos e outros ônus de manutenção das URL's presentes nos registros do catálogo InterCat - Internet Catalog Project - projeto da OCLC voltado para a investigação e o estabelecimento de práticas e normas de catalogação para recursos Internet que rodou de 1993 a 1996 -, e assegurar o acesso de longo prazo aos recursos digitais referenciados por ele (Dack, 2001).

Um dos objetivos subjacentes ao esquema PURL é contornar a atual falta de consenso e de progresso nas questões de nomes na Internet e, ao mesmo tempo, estabelecer práticas concernentes ao uso de identificadores persistentes em sistemas bibliográficos. Nessa direção, o sistema PURL preenche algumas das especificações do URN, e pode ser considerado uma solução intermediária aguardando o momento em que o URN se torne uma parte efetiva da arquitetura de informação da Web. A OCLC tem como perspectiva, para um futuro oportuno, poder traduzir mecanicamente

o sistema PURL para a forma URN (Weibel; Jul; Shafer, 1995).

Resolução de PURL

O PURL é, funcionalmente, um URL cujo endereço do servidor é o nome do serviço de resolução PURL. Ou seja, ao invés de apontar diretamente para a localização do recurso, o PURL aponta para um serviço de resolução intermediário, o qual, por sua vez, consulta uma base de dados que armazena a localização atualizada do recurso. Em termos de tecnologia *web*, trata-se simplesmente de um redirecionamento padrão HTTP - Hypertext Transfer Protocol - no qual não se exige uso de outros protocolos, nem se exige modificações no software cliente (Shafer et al., 1996).

Dessa forma, o PURL minimiza o problema da transitoriedade da localização do URL, na medida em que ele é sempre traduzido pelo URL atual do recurso. Uma vez que o PURL de um recurso é registrado na base de dados, ele pode ser usado para substituir o URL atual do objeto em diversas instâncias onde é desejável maior estabilidade, tais como registros bibliográficos em catálogos, ou em outros tipos de metadados que estabelecem *link* com o recurso. Nesta situação, se a URL do recurso mudar por algum motivo, a única ação de manutenção necessária é a atualização na base de dados do serviço de resolução. O Projeto InterCat demonstrou como o PURL pode ser usado em catálogos on-line, incluindo os PURL's no campo 856 do MARC (Weibel; Jul; Shafer, 1995; Payette, 1998).

Estrutura

Como é de se esperar, o PURL se assemelha estruturalmente ao URL, dado que PURL's são verdadeiramente URL's. O PURL é composto por: 1) protocolo; 2) endereço do serviço de resolução; e 3) nome do recurso. O exemplo a seguir usa o protocolo de acesso HTTP para conectar o Serviço de Resolução PURL, que, por sua vez, resolve o nome.

<http://purl.oclc.org/OCLC/PURL/FAQ>

onde "http" é protocolo; "purl.oclc.org" é o endereço do resolvedor - endereço IP ou nome de domínio do resolvedor PURL - e "OCLC/PURL/FAQ" é o nome do recurso.

Disponibilidade

Embora o Sistema PURL seja operado e mantido pela OCLC, o modelo PURL presta-se para distribuição na rede, por meio de servidores operados por organizações que tenham o compromisso de manter esquemas de nomes persistentes. A OCLC distribui gratuitamente o código-fonte do software PURL para as organizações que queiram estabelecer seus próprios serviços de resolução. Devido a isto, um número significativo de instituições já opera servidores com tecnologia PURL. A lista completa pode ser consultada na página da OCLC/PURL (<http://purl.oclc.org/>).

Handle System

O *Handle System* (<http://www.handle.net>) é um sistema distribuído de computadores concebido para assinalar, armazenar, administrar e resolver identificadores ou nomes persistentes de objetos digitais conhecidos como *handles*. Esses nomes podem ser rapidamente resolvidos em informação necessária para localizar e acessar os objetos aos que se referem, num ambiente de rede tal como a Internet (Lannom, 2000). As informações de localização e acesso podem sofrer mudanças, refletindo o estado atual do recurso, sem que haja, entretanto, alteração no seu *handle*. Isto permite que a identificação do item transcenda as mudanças de localização, de propriedade e de outras informações transitórias, garantindo a característica desejável de persistência.

O *Handle System* é caracterizado como um sistema de informação de grande amplitude, projetado para alcançar interoperabilidade global através de uma rede hierarquicamente distribuída de servidores. O objetivo inicial do Sistema era estabelecer uma infraestrutura básica no domínio das bibliotecas digitais e das publicações eletrônicas. Entretanto, o *Handle System* criou uma arquitetura de tal forma abrangente, que aplicações de propósitos diversos podem ser hoje desenvolvidas tomando-a como base (Payette, 1998; Lannom, 2000; Dack, 2001). O sistema constitui-se também numa implementação do URN - Uniform Resource Names -, posto que seu esquema de nomes, resolução e estrutura de registros, estão em conformidade com as especificações RFC 1737, *Functional Requirements for Uniform Resource Names* (Sollins, 1994), já descrito anteriormente neste trabalho.

O *Handle System* foi desenvolvido pelo *Corporate for National Research Initiatives (CNRI)* (<http://www.cnri.reston.va.us/>) - e teve sua origem no projeto *Networked Computer Science Technical Report Library (NCSTRL)* (<http://www.ncstrl.org/>). Parte desse projeto estava voltada para o estabelecimento de uma arquitetura de apoio ao desenvolvimento de uma biblioteca digital aberta e distribuída (Dack, 2001). Somando-se a esta aplicação inicial, o sistema foi implementado também, entre outras experiências, pela *Library of Congress* no contexto do *National Digital Library Program* (<http://memory.loc.gov/ammem/dli2/html/lcndlp.htm>), e no *US Copyright Office, CORDS* (<http://www.copyright.gov/cords/index.html>), onde *handles* são assinalados para os objetos digitais quando lá são depositados. Uma aplicação do *Handle System* relativamente conhecida no Brasil é a que é adotada pelo software *Dspace* (<http://www.dspace.org/>). Entretanto, o exemplo mais conhecido em escala global é o *DOI System* (<http://www.doi.org/>), que é uma aplicação do *Handle System* voltada para gestão de *copyright* de objetos digitais. Este sistema será analisado em detalhes na próxima seção deste artigo.

Estrutura do Handle System

O *Handle System* está além de um simples esquema de atribuição de nomes. O projeto incorpora um conjunto expansível de servidores de *handles*, que se comunicam entre si segundo um conjunto de protocolos abertos e um sistema de servidores *cache* intermediários, que reduz a carga sobre a rede, salvando dados sobre *handles* recentemente usados, incluindo os de uso freqüente. Para tal, o *Handle System* se estrutura num modelo hierárquico de serviço que possui dois níveis: o mais alto consiste em um único serviço - o *Global Handle Registry*; no outro nível estão os demais serviços de identificação, que são genericamente conhecidos como *Servidores Locais de Handles (LHS - Local Handle Services)*. O papel de cada um desses atores é resumido a seguir:

a) **Autoridades Nomeadoras (Naming Authorities)**: São unidades administrativas autorizadas a criar e a manter identificadores (*handles*). O *Global Handle Registry*, operado pela *CNRI*, configura-se na maior instância hierárquica entre as autoridades nomeadoras do sistema. A criação das demais autoridades é responsabilidade atribuída aos diversos níveis da hierarquia. Cada autoridade nomeadora estabelece um domínio para atribuição de nomes - chamado de *namespace*. No âmbito de um

namespace, cada nome atribuído a um recurso é único; cada *namespace* local é vinculado a uma única autoridade nomeadora; cada autoridade nomeadora tem a garantia de ser globalmente única dentro do contexto do *Handle System*. Este mecanismo garante duas coisas importantes num esquema de identificadores: a singularidade de um nome atribuído no contexto do *Handle System* e o correto roteamento das requisições de resolução.

O *Global Handle Registry* é responsável por conhecer a localização e o *namespace* de todos os serviços locais de *handles*; por sua vez, cada um desses serviços locais sabe como acessar o *Global Handle Registry*. Isto permite que uma consulta possa ser submetida ao *Handle System* em qualquer ponto do sistema e possa ser roteada para o servidor que conhece a resposta. Qualquer servidor de *handle* no mundo - ou navegador habilitado a reconhecer *handles* - tem acesso a todos as autoridades nomeadoras e, conseqüentemente, a todos os nomes de uma maneira transparente para o usuário. A autoridade nomeadora não é vinculada a um endereço internet específico (*Library of Congress, 1997*).

b) Servidores de nomes (*handle servers*)

- armazena nomes e disponibiliza um serviço para resolvê-los quando solicitado por um usuário. Existe um único *Global Handle Server* ou *Global Handle Registry* e muitos servidores locais hierarquicamente associados a ele e mantidos por autoridades nomeadoras locais. O *Global Handle Registry* armazena e resolve *handles* para todas as autoridades nomeadoras e servidores locais, e fornece ao *software-cliente* informações sobre qual(is) servidor(es) local(ais) podem operar uma requisição.

Sintaxe do Handle

Dentro do contexto de um *namespace*, um *handle* consiste em duas partes: 1) "Prefixo", que identifica a autoridade nomeadora, a unidade administrativa autorizada a criar e manter nomes; 2) "Sufixo", nome do recurso, único dentro do domínio da autoridade nomeadora. A autoridade e o nome são separados pelo caractere ASCII *"/"*. Formalmente um *handle* pode ser definido como:

<Handle> := <Handle Name Authority>/<Handle Local Name>

É importante notar que o nome do recurso não possui nenhuma sintaxe pré-estabelecida, sendo de responsabilidade da autoridade local, com cujas regras específicas deve estar em conformidade.

Resolução do Handle System

O *Handle System* permite que *handles* possam ser resolvidos de uma forma distribuída, usando-se *softwares*-clientes dedicados ou *softwares*-clientes comuns, como são os navegadores (*web browsers*). Para este último caso, é imprescindível a utilização de *plug-ins* ou servidores *proxy*, que fazem a intermediação entre o protocolo HTTP usado pelos navegadores e os protocolos de resolução de *handles* - Handle System Protocol. A CNRI opera um servidor *proxy* em <http://hdl.handle.net/> e a Library of Congress roda outro no endereço <http://hdl.loc.gov/>. Seja qual for o servidor *proxy*, ele pode resolver qualquer *handle* que tenha um URL correspondente. Vejamos o seguinte exemplo: o *handle* "10.1045/march2003-marcondes" identifica o artigo "The SciELO Brazilian Scientific Journal Gateway and Open Archives" publicado no periódico D-Lib Magazine, que tem como prefixo "10.1045". Atualmente ele aponta para o URL

<http://www.dlib.org/dlib/march03/marcondes/03marcondes.html>

Prefixando o *handle* "10.1045/march2003-marcondes", com "http://" e o nome de domínio Internet de um servidor *proxy* de *handle* qualquer, ele pode ser resolvido usando os *softwares* navegadores atuais. Assim sendo, os URLs abaixo relacionados recuperam o mesmo artigo:

<http://hdl.handle.net/10.1045/march2003-marcondes>

<http://hdl.loc.gov/10.1045/march2003-marcondes>

<http://purl.handle.net/10.1045/march2003-marcondes>

Softwares do Handle System

O *software* do *Handle System* - escrito em linguagem Java - está disponível publicamente para *download* na página *web* do CNRI, para pesquisa e uso experimental. Os desenvolvedores de aplicações voltadas para repositórios digitais, dispõem de uma ampla biblioteca de *softwares* que permitem aos seus sistemas comunicarem-se diretamente com o *Handle System*. A biblioteca inclui *softwares* para o estabelecimento de estruturas de atribuição de nomes: *software* para serviços locais, *software*-cliente e ferramentas de gestão, servidores *cache*, ferramentas

para a criação e administração de *handles* e autoridades nomeadoras, servidores *proxy* que permitem aos navegadores *web* resolver nomes. O CNRI mantém ainda um servidor de teste que permite ensaios de resolução.

Digital Object identifier

O *Digital Object Identifier* (DOI) (<http://www.doi.org>) é uma aplicação específica do *Handle System*, voltada para a identificação persistente de recursos digitais aos quais possam ser atribuídos direitos de propriedade intelectual, bem como para o intercâmbio de informações sobre essas propriedades em um ambiente de rede. Além de identificar, o DOI associa aos objetos digitais dados estruturados - informações bibliográficas e comerciais atualizáveis. Para tanto, o DOI estabelece uma infra-estrutura ampla, cuja perspectiva é ligar os usuários aos fornecedores de conteúdo, dentro de um escopo que considera sempre a facilitação das práticas de comércio eletrônico de conteúdos e a viabilidade da gestão automática de *copyright*. Dessa forma, o escopo do DOI se estende para além de ser um simples identificador, abrangendo também os desafios de gerenciar o comércio eletrônico, as questões ligadas ao *copyright* para objetos digitais publicados na Internet, o controle de transações e ainda a comunicação entre os editores e seus clientes (Payette, 1998). A idéia central dos criadores do DOI foi desenvolver um identificador padronizado de extensão industrial, que pudesse ser assinalado a um trabalho, ou a manifestações de um trabalho, no momento da sua criação, consubstanciando-se numa resposta à crescente preocupação dos produtores de conteúdo quanto à proteção de direitos em redes eletrônicas abertas (Dack, 2001).

O DOI, enquanto uma aplicação do *Handle System*, se constitui também numa implementação dos conceitos circunscritos pelo *Uniform Resource Name* (URN) e pelo *Uniform Resource Identifier* (URI). No entanto, a metodologia DOI apresenta uma característica importante: ela agrega uma infra-estrutura técnica e social a esses conceitos teóricos, viabilizando-a pela adoção de um Modelo de Dados - o qual será analisado mais adiante.

O DOI foi oficialmente lançado na Frankfurt Book Fair em 1997. O seu desenvolvimento é de responsabilidade da International DOI Foundation (IDF),

uma instituição sem fins lucrativos fundada por iniciativa da Association of American Publishers (AAP), cujo corpo de membros é composto por representantes de grandes editores, publicadores e empresas de software, bem como organizações que representam os interesses dos detentores de direitos autorais, como por exemplo, a International Publishers Association (Dack, 2001). Apesar do seu viés comercial, o DOI pode ser aplicado a qualquer forma de propriedade intelectual que se manifeste em um meio digital. Pode ser usado para identificar recursos que são objetos de transações de propriedade intelectual, envolvendo dinheiro ou não. O DOI identifica também recursos livres. Ele pode ser usado para identificar textos, áudio, vídeos, imagens, software, etc. Entretanto ele difere de outros identificadores de propriedade intelectual, tal como o ISBN, amplamente utilizado no domínio bibliográfico, por estar associado com serviços e ser diretamente acionável - ou seja, "clicável" - num ambiente de rede.

Para alcançar os seus objetivos, o DOI vem desenvolvendo um sistema completo apoiado em três componentes básicos: a) uma sintaxe e regras de atribuição do identificador DOI; b) um sistema de resolução; e c) uma estrutura de metadados que, além de identificar sem ambigüidades o item representado pelo DOI, confere um grau de inteligência aos processos interoperáveis.

Como o Sistema DOI funciona

Sintaxe DOI

A sintaxe DOI é definida pela norma ANSI/NISO Z39.84-2000: *Syntax for the Digital Object Identifier*, a qual estabelece, resumidamente, as seguintes regras: o DOI é constituído de dois componentes, "prefixo" e "sufixo". Todo prefixo começa pelo Código de Diretório, que atualmente é sempre "10", embora códigos adicionais possam ser assinalados no futuro. O Código de Diretório é seguido de um número atribuído pela Agência de Registro (AR) a grupos tais como editores, publicadores, detentores de direitos autorais, etc. - aos quais, para facilitar a leitura, chamaremos aqui de gestores de conteúdo. Estes gestores podem optar por requisitar um prefixo para cada linha de produto que eles controlam, ou podem usar um prefixo único (International DOI Foundation, 2006). O segundo elemento, o sufixo, é uma seqüência de caracteres (*string*) assinalada pelo gestor de conteúdo para um recurso específico, objeto da identificação.

Identificadores já comumente usados em instâncias específicas - como, por exemplo, o ISBN, usado para identificar livros -, ou esquemas proprietários, usados no contexto de uma determinada organização, podem ser incorporados ao sufixo do DOI, compondo a identificação do recurso e agregando a estes identificadores as características do DOI. A sintaxe do DOI é a seguinte:

<Código do Diretório>.<Código de Registro>/
<Seqüência Sufixo DOI>

Os sufixos, no domínio de um gestor de conteúdos, podem ser assinalados para recursos de qualquer tamanho ou granularidade - livros, artigos, resumos, música, mapas, etc. -, ou para qualquer tipo de arquivo - texto, áudio, vídeo, imagem, software, etc. Fica por conta do gestor decidir o nível ou granularidade da identificação, baseado na natureza dos recursos a serem distribuídos e/ou vendidos eletronicamente (International DOI Foundation, 2006).

Exemplo:

O DOI "10.1016/S1057-2317(03)00016-X" tem como prefixo "10.1016", que identifica o gestor/produtor/distribuidor do conteúdo; enquanto o sufixo, dado pela seqüência "S1057-2317(03)00016-X" identifica um objeto específico e único, ou seja, neste exemplo, o artigo: MARCONDES, Carlos Henrique; SAYÃO, Luis Fernando. Brazilian Digital Library of Theses and Dissertations. *The International Information & Library Review*, v.35, n.2-4, p.265-279, June-December 2003.

Resolução

O DOI utiliza uma tecnologia Internet que permite ao identificador ter a característica de ser "clicável". Este processo é realizado pelo Handle System. É importante assinalar que, na implementação inicial, o DOI era resolvido, via de regra, para uma única versão, a versão *default*, geralmente publicada no *website* do editor ou de uma organização licenciada, correspondendo a um único URL. Entretanto, com o desenvolvimento pleno do Sistema, associado a uma estrutura de metadados, vislumbrou-se a possibilidade da resolução múltipla, que possibilita a resolução em várias versões, abrindo possibilidades de provimento de serviços integrados, com maior grau de customização e sensíveis ao contexto do usuário. (Paskin, 2003).

Exemplo:

Clicando o link [http://dx.doi.org/10.1016/S1057-2317\(03\)00016-X](http://dx.doi.org/10.1016/S1057-2317(03)00016-X) o usuário é conduzido para o portal da Science Direct, onde estão disponíveis informações sobre o artigo identificado pelo DOI "10.1016/S1057-2317(03)00016-X", incluindo o acesso ao texto completo para usuários autorizados ou o acesso por *pay-per-view*.

Metadados DOI

Para alcançar as funcionalidades planejadas pelos seus criadores, o DOI deve estar associado sem ambigüidades ao recurso que ele representa, além de ser capaz de ser resolvido para um número variado de diferentes localizações, ou seja, capaz de resolução múltipla. Para tal, é necessário que haja um nível mínimo de metadados estruturados, contendo informações bibliográficas e comerciais disponíveis publicamente e associadas ao recurso. Estes metadados servem para assegurar ao usuário que o recurso encontrado é realmente o que estava sendo procurado; servem ainda, para dar conhecimento sobre versões e manifestações de um objeto, sem a necessidade de consultá-lo, viabilizando uma seleção inteligente entre múltiplas localizações do mesmo recurso. Um esquema estruturado de metadados facilitará também a introdução de soluções automatizadas nos processos de seleção e nas transações comerciais, além de um alto grau de interoperabilidade entre diferentes serviços e sistemas (Dack, 2001).

Para ancorar os serviços mais sofisticados - que estão além da sua função de identificador - e assegurar a interoperabilidade entre diferentes sistemas, o DOI define um Modelo de Dados baseado no Projeto <indec> - *Interoperability of Data in E-Commerce Systems* (<http://www.indec.org/>) -, o qual provê um sistema contextual de metadados que visa dar suporte à interoperabilidade dos múltiplos esquemas de metadados no âmbito do comércio eletrônico.

O Modelo de Dados DOI tem como objetivo principal promover a interoperabilidade no domínio da rede de usuários DOI, por meio do estabelecimento de estratégias de "compatibilidade semântica" entre diferentes Agências Registradoras (RA), assegurando a viabilidade de meios de intercâmbio eficiente e extensível para transporte de metadados entre as RA's

(*International DOI Foundation*, 2006). O Modelo de Dados do Sistema DOI consiste de um dicionário de dados interoperável e uma estrutura subjacente para aplicá-lo. O dicionário de dados é construído tendo como fundamento uma ontologia - a ontologia IDF - e tem como finalidade assegurar máxima interoperabilidade com elementos de outros esquemas de metadados. Em conjunto, essa metodologia proporciona ferramentas para definir o que um nome DOI específica - efetivado através do uso do dicionário de dados -, e como os nomes DOI se relacionam - efetivado através de um mecanismo de agrupamento chamado "Perfis de Aplicação" -, que associam nomes DOI com propriedades comuns previamente definidas, significando que certos nomes DOI se comportem todos de forma previsível em uma aplicação, no contexto de um serviço. Este conjunto de dispositivos viabiliza a interoperabilidade semântica, ou seja, permite que informações originadas em um contexto possam ser usadas em outro, da maneira a mais automatizada possível. Por exemplo: muitas RA's e provedores de recursos - como os editores de periódicos - necessitam que DOI's emitidos por outras RA's estejam disponíveis para inclusão em serviços disponibilizados por eles. Esta característica de interoperabilidade é o principal benefício do Sistema DOI (DOI System Overview, 2006).

Para dar sustentação à política de metadados DOI, o Modelo de Dados lança mão de três ferramentas interoperáveis: 1) *Kernel Metadata Declaration* (KMD) - conjunto mínimo de informações, expressas por oito elementos de metadados descritivos que asseguram que DOI's possam ser descobertos e tenham ambigüidades eliminadas de forma coerente num ambiente de múltiplos serviços e múltiplos Perfis de Aplicação; 2) *Resource Metadata Declaration* (RMD) - esquema voltado especificamente para o intercâmbio de metadados entre AR's num determinado domínio, por exemplo, periódicos (*Journal-RMD*) e e-books. O RMD usa a estrutura genérica de metadados composta de dez elementos, projetada para incorporar metadados para todos os tipos de recursos; 3) *Indec Data Dictionary* (IDD) - repositório para registro de todos os elementos e valores permitidos usados no KMD e no RMD. O IDD permite que uma definição e uma ontologia estruturada para todos os elementos de metadados estejam disponíveis para as AR's. Oferece as condições para o mapeamento necessário à integração e para as transformações exigidas para o intercâmbio de dados entre as Agências Registradoras. Por exemplo, se uma

AR deseja consolidar metadados fornecidos por várias outras Agências tendo em vista um serviço específico, o IDD irá prover os dados de mapeamento necessários para que a AR viabilize essa tarefa (International DOI Foundation, 2006).

Conformidade com os padrões de interoperabilidade

O DOI foi projetado tendo como perspectiva a máxima conformidade com as normas e padrões correntes voltados para a interoperabilidade. Em consequência, observam-se as seguintes características: o Sistema DOI como um todo está sendo objeto de normalização pela ISO - ISO TC46/SC9; o DOI está em conformidade com os conceitos URN e URI, estabelecido pelo RFC 2141 URN Syntax (Moats, 1997); o Sistema de Metadados DOI tem como fundamento o <indec> Framework, sendo expressos em RDF; a sintaxe do DOI está registrada como uma norma ANSI/NISO - Z39.84-2000 *Syntax for the Object Identifier*.

CROSSREF

As referências constituem um núcleo de vital importância nos periódicos científicos. É consenso que a citação num texto acadêmico, expresso por meio de referências bibliográficas, é um dos fundamentos básicos do sistema corrente de comunicação científica. Através da citação, o autor realiza um vínculo explícito entre a sua pesquisa e outros trabalhos, estabelecendo um contexto multifacetado - histórico, social, metodológico, etc. - para sua obra. Nada mais interessante para o leitor do que ter visão desse contexto por meio de um ou dos cliques imediatamente acionáveis (Brand, 2001).

Com o surgimento dos periódicos eletrônicos, tornou-se possível o desenvolvimento de mecanismos que permitem o acesso imediato, via *links*, a trabalhos referenciados num artigo. Esta facilidade é hoje algo tão importante para os usuários, que se transformou em um imperativo econômico para os editores científicos: hoje, ter disponível nas publicações eletrônicas "*links* referenciais", tornou-se uma característica essencial. Entretanto, os primeiros sistemas de *links* referenciais tinham um escopo bastante reduzido, pois eram focados em áreas específicas e circunscritos a sistemas proprietários de periódicos

(Pentz, 2001). Reconhecendo a importância dos *links* referenciais para os seus usuários e para os seus negócios, conscientes das limitações e ineficiência dos sistemas existentes, e, sobretudo, entendendo que a economia dos periódicos científicos - baseado em subscrição - se desloca para uma economia de venda de artigos, os editores científicos, numa iniciativa incomum de cooperação, estabeleceram as bases para um serviço colaborativo de *links* referenciais. Foi neste contexto que surgiu o CrossRef, no ano 2000 (Brand, 2001; Pentz, 2001).

CrossRef (<http://www.crossref.org>) é uma associação cooperativa independente, fundada e dirigida por editores científicos, tendo como missão ligar usuários a conteúdos de pesquisas, através de tecnologias colaborativas e acordos coletivos entre editores - muitas vezes competidores entre si - acadêmicos e profissionais (Brand, 2003). O objetivo inicial do CrossRef era se tornar nada menos do que a mais completa espinha dorsal (*backbone*) de *links* referenciais, cobrindo todas as publicações acadêmicas eletrônicas (Brand, 2001) e estabelecendo uma integração virtual via *links*, na qual o conteúdo propriamente dito permanece no *website* do editor ou de seus licenciados. O sistema é operado pela PILA - Publishers International Linking Association Inc, cujo *Board of Directors* reúne representantes das mais importantes casas editoras científicas do planeta, tais como Academic Press, Elsevier, IEE, Nature, Oxford University Press, Springer Verlag, Science, John Wiley & Sons.

O CrossRef não é um produto, é um processo (Brand 2001), cuja característica marcante é oferecer uma infra-estrutura que sustente um sistema de referências cruzadas via *links*. Isto permite que o usuário, ao clicar numa referência citada em um periódico publicado pelo editor A, seja conduzido diretamente ao conteúdo referenciado, publicado num periódico do editor B, ressalvadas as práticas de controle de acesso do editor B e/ou as modalidades de acesso oferecidas ao usuário - subscrição, acesso via IP, *pay-per-view*, acesso livre etc.

Como funciona a citação cruzada via links referenciais

O sistema CrossRef é sustentado por padrões abertos e, portanto, o mecanismo para *links* entre

referências, estabelecido cooperativamente pelos editores científicos, tem como base o consenso em torno do DOI como identificador para os itens de suas coleções. O Sistema constitui, essencialmente, uma base de dados *CrossRef Metadata Database* (MDDDB) - onde os editores associados depositam os DOI's e os URL's correspondentes, além de um conjunto mínimo de metadados - título do periódico, volume, número, ISSN, paginação e autor principal - de artigos científicos publicados por eles. Esta submissão é feita em formato XML, segundo a *Document Type Definition* (DTD), disponível no *website* do CrossRef (<http://www.crossref.org>). Como parte do processo, os DOI's e os URL's dos artigos são registrados no Diretório Central DOI, operado pela DOI Foundation (<http://www.doi.org>) (Atkins et al., 2000).

Para os editores associados, o CrossRef oferece três serviços básicos: o depósito de metadados de artigos na base de dados do CrossRef; a submissão de referências de artigos com o propósito de obter seus DOI's; e a criação de *links* usando esses DOI's.

Por outro lado, os editores associados podem dirigir consultas à base de dados MDDDB, através do seu *front-end*, o *Reference Resolver* (RR), para recuperar DOI's de citações contidas nas suas publicações. Essas consultas têm como argumento os metadados das referências citadas, formatadas de acordo com um protocolo próprio do CrossRef, cujas regras estão disponíveis no *site* do sistema. O RR permite a recuperação de DOI's, que vão viabilizar a criação de *links* CrossRef associados às citações, os quais, quando clicados, serão resolvidos pelo Diretório DOI nas URL's correspondentes, previamente depositadas pelos editores (Brand, 2001).

Suponhamos que um editor encontre a referência abaixo, citada em uma de suas publicações: MARCONDES, C.H.; Sayão L.F. 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.

Ele pode então consultar a base de dados CrossRef, usando os elementos de metadados da citação para recuperar o DOI: 10.1045/march2003-marcondes

Isto concluído, esse editor pode incluir um *link* DOI, baseado no CrossRef, na citação que aparece na sua publicação, na forma de um *link* URL acionável: <<http://dx.doi.org/10.1045/march2003-marcondes>>.

Quando um usuário clica neste link, o servidor *proxy* DOI, que é o alvo de todos os *links* no *namespace* CrossRef, passará o DOI para o servidor de *handle* DOI. O servidor de *handle*, por sua vez, resolve o DOI em um URL para o qual o usuário é redirecionado.

Outro serviço que está sendo crescentemente utilizado é a consulta à base CrossRef, tendo como argumento o DOI, com o objetivo de recuperar os metadados bibliográficos depositados pelo editor (O'Neill, 2003). Esta funcionalidade tem grande importância no oferecimento de serviços sensíveis ao contexto, como será visto na próxima seção.

OPENURL -Integrando os serviços de links referenciais

Quando um usuário clica sobre um *link* presente num recurso informacional, como por exemplo, uma citação num periódico eletrônico, ele é redirecionado sempre para a mesma versão do conteúdo, a versão *default*, geralmente publicada no *website* do editor. Isto acontece porque os *links* convencionais não levam em consideração a identidade de cada usuário, seu contexto institucional, suas preferências e direitos em termos de acesso e de serviços disponíveis para os usuários, conduzindo-os todos, indistintamente, para a mesma fonte. Esta é uma resolução fechada e não sensível ao contexto, posto que esta versão pode não ser a única disponível nem a mais conveniente para um determinado usuário (van de Sompel; Beit-Arie, 2001). Idealmente, se mais de um fornecedor oferece serviço de cópias para um mesmo artigo eletrônico, o *link* de uma citação que leva ao texto completo desse artigo, deveria conduzir o usuário para o serviço que oferece a cópia mais conveniente para ele. Por exemplo, a instituição do usuário pode não ser assinante direta do periódico, mas pode ter direito a oferecer ao usuário acesso via algum portal agregador - como é o caso do Portal Capes - ou possuir assinatura em papel do periódico. Seria mais adequado que a biblioteca apresentasse um conjunto de opções que superassem aquela única disponibilizada pelo *website* do editor, além de outros serviços adicionais ligados ao recurso (O'Neil, 2003). Para realizar esta função, um *link* deve ser capaz de: a) empacotar metadados bibliográficos e identificadores que descrevam e identifiquem o objeto informacional que ele está referenciando; b) enviar o pacote para um servidor de resolução de *links*, ou um resolvidor de *links*, ou simplesmente um servidor de *links*, capaz de interpretar esses metadados.

Se o servidor de *links* for informado sobre o contexto do usuário, ele será capaz de considerar a identidade do usuário quando da resolução dos metadados, direcionando-o para um serviço cujo acesso lhe seja o mais conveniente e/ou apresentando um menu de opções de serviços. Isto configura uma solução aberta, sensível ao contexto. A seleção da melhor cópia, ou da “cópia apropriada” - como é chamada pelos autores da área -, é baseada nas preferências do usuário e nas políticas de informação da sua organização, considerando a localização da cópia, seu custo, os acordos comerciais com fornecedores, com serviços agregadores e com os estoques de informação da instituição (Caplan; Arms, 1999; Beit-Arie et al., 2001; van de Sompel; Beit-Arie, 2001).

O cerne da tecnologia sensível ao contexto, ou seja, tecnologia *open reference linking*, é a infraestrutura OpenURL. Ela é definida pela norma *OpenURL da National Information Standards Organization (NISO)*, e foi inicialmente desenvolvida por Van Sompel e Oren Beit-Arie (van de Sompel; Beit-Arie, 2001). As especificações estabelecidas pelo OpenURL têm como principal objetivo estender a vinculação por *links* entre recursos de serviços secundários de informação *on-line* - registros de uma base de dados, por exemplo - e o conteúdo de periódicos eletrônicos. A essência da norma é tornar possível a criação de *links* para um determinado recurso informacional, sem a necessidade de conhecer nenhum URL específico, ou de conhecer as possíveis opções de acesso a este recurso que estão disponíveis para uma particular instituição ou grupo de usuários (van de Sompel, 1999). O OpenURL desempenha esta função transferindo a responsabilidade de identificação da localização do recurso para o serviço de informação local - na maioria dos casos, a biblioteca - e, dessa forma, tornando viável a possibilidade da escolha da cópia apropriada para um determinado usuário. (Caplan; Arms, 1999; Langston; Tyler, 2004).

Com um grau a mais de rigor, a norma OpenURL define um protocolo para interoperabilidade entre um recurso de informação e um componente de serviço, denominado servidor de *links*; este, oferece serviços de localização sensíveis ao contexto, através da interpretação de metadados bibliográficos, codificados segundo uma sintaxe própria, que descrevem uma publicação. O servidor de *links* que esteja em conformidade com a norma OpenURL, define o contexto

do usuário de forma a viabilizar serviços que estejam aderentes às políticas, às condições de acesso e às coleções da instituição desse usuário.

A norma OpenURL especifica um formato padronizado para transportar metadados bibliográficos de objetos informacionais entre serviços de informação (van de Sompel; Beit-Arie, 2001). Ou seja, define uma sintaxe para a criação de um pacote transportável, via *web*, de metadados e/ou identificadores que referenciam um recurso informacional, tipicamente um recurso bibliográfico. Essa sintaxe permite a formatação de pacotes na forma de um URL, por meio da codificação dos metadados do recurso - título, autor, título de periódico, etc. - ou através do uso de um identificador global do recurso, como o DOI, ou ainda, através de uma combinação dos dois métodos. Desta forma, o OpenURL é, funcionalmente, um URL acionável, cujo alvo é um serviço de resolução OpenURL - servidor de *links* OpenURL - que oferece serviços sensíveis ao contexto, baseado na interpretação dos metadados (Powell, 2001).

Entretanto, para que isto aconteça, é necessário que o servidor de *links* OpenURL esteja configurado pela biblioteca local para refletir os seus acervos e direitos de acesso a outros serviços. Um OpenURL enviado por um serviço de informação a um servidor OpenURL, terá seus dados comparados aos serviços disponíveis para o usuário e serão, finalmente, exibidos em forma de menu, para a escolha deste usuário (Langston; Tyler, 2004).

A estrutura de um OpenURL é composta por duas partes: um “URLBase” e uma “Consulta”. O URLBase identifica o servidor OpenURL que fornecerá o serviço sensível ao contexto para um particular OpenURL. O URLBase é específico para um usuário particular que aciona um OpenURL, posto que este identifica o resolvedor OpenURL preferido pelo usuário, o qual, na maioria dos casos, será o resolvedor oferecido pela instituição do usuário. Os serviços de informação que incorporam OpenURL nas suas interfaces *web* - no resultado de consultas às bases de dados, por exemplo - devem possuir mecanismos que associem uma URLBase a cada usuário-final. Uma das maneiras de fazer isto é registrar a informação sobre o URLBase preferida na forma de *cookie*, no próprio navegador do usuário (mais informações em: http://www.exlibris-usa.com/sfx_cookiepusher.htm); a outra, é armazenar a URLBase junto com outras preferências

do usuário (Powell, 2001). A “consulta” pode ser composta por uma ou mais “descrições”. Cada descrição compreende os atributos de metadados e valores que compõem a citação do recurso. Detalhes sobre a sintaxe do OpenURL estão na especificação *OpenURL Syntax Description* (van de Sompel; Hochstenbach; Beit-Aire, 2000). Segue-se o exemplo de um OpenURL.

```
http://resolver.ukoln.ac.uk/openresolver/
?sid=ukoln:ariadne&genre=article
```

```
&title=Information%20gateways:%20collaboration%20on%20content
```

```
&title=Online%20Information%20Review&issn=1468-4527&volume=24
```

```
&page=40&epage=45&artnum=1&aualast=Heery&aufirst=Rachel
```

O URLBase é: <http://resolver.ukoln.ac.uk/openresolver>, que é o URL do UKOLN Open Resolver Demonstrator Service. Os dados seguintes formam a “consulta”, em que se identifica a descrição de um artigo intitulado *Information Gateways*, escrito por Rachel Heery em colaboração e publicado no periódico *Online Information Review*, volume 24 (Powell, 2001).

O software que resolve o OpenURL extrai os metadados embutidos na “consulta” OpenURL e os utiliza para montar URL’s que apontam para uma das versões do recurso e/ou consultam outros serviços de informação compatíveis com esta abordagem. Por exemplo:

- o DOI pode ser resolvido no URL de uma das versões on-line do artigo;

- as palavras principais do título podem ser usadas como base para a formulação de consulta às máquinas de busca - Google, Yahoo, etc.;

- o título do periódico ou o ISSN pode ser usado para busca no catálogo on-line da biblioteca; no caso de livros; pode-se usar o ISBN para consultas on-line às bibliotecas;

- o nome do autor pode ser usado para busca em bases de dados de currículos, como o Diretório Lattes, por exemplo.

Os mais importantes serviços secundários de informação e muitos periódicos eletrônicos já são compatíveis com as especificações OpenURL. Isto significa que eles estão habilitados a enviar requisições OpenURL para quaisquer servidores que estejam em conformidade com o padrão OpenURL (Langton, 2004).

Um número crescente de empresas passou a oferecer softwares resolvidores de OpenURL. SFX, da Exlibris, foi o primeiro produto do mercado e tornou-se uma referência (<http://www.exlibrisgroup.com/sfx.htm>). A UKOLN - UK Office for Library and Information Networking - disponibiliza um script CGI de fonte aberta para resolução de OpenURL’s, o *OpenResolver*, o qual pode ser baixado livremente em <<ftp://ftp.ukoln.ac.uk/metadata/tools/openresolver/>>. Uma lista de produtos OpenURL pode ser consultada em <<http://www.loc.gov/catdir/lcpaig/openurl.html>>. Para mais informações sobre a norma OpenURL consulte <<http://library.caltech.edu/openurl/>>.

Juntando tudo em torno da interoperabilidade:

DOI, Handle System, CrossRef e OpenURL

O DOI, o *Handle System*, o *CrossRef* e o *OpenURL* possuem um grande potencial para trabalhar em conjunto na otimização da interoperabilidade entre serviços de informação, através da integração de sistemas de *links* referenciais no ambiente, por exemplo, de uma biblioteca. Com intuito de compreender melhor como isso pode se concretizar, vamos revisar rapidamente alguns conceitos já vistos anteriormente:

- 1) O *OpenURL* é um mecanismo para o transporte de metadados e identificadores que descrevem uma publicação, tendo como perspectiva tornar viáveis serviços sensíveis ao contexto. 2) um resolvidor de *links* local é um servidor de *links* que funciona dentro de um contexto institucional. Ele pode interpretar *inputs* formatados segundo as especificações OpenURL, tomando em conta os acervos locais e privilégios de acesso da instituição - geralmente a biblioteca - e, dessa forma, ele pode apresentar *links* para recursos apropriados. 3) *CrossRef* é um sistema de links referenciais, construído cooperativamente por editores científicos, voltado para identificação persistente de conteúdos acadêmicos e de citação cruzada entre publicações de editores distintos. Ele utiliza o DOI para estabelecer *links* para o texto completo de citações. Na base de dados do *CrossRef* são depositados pelos editores o DOI e os metadados de suas publicações. 4) *Handle System* resolve um particular DOI em um URL.

A infra-estrutura de *links* estabelecida pelos serviços de *links* referenciais relacionados acima confere inteligência ao sistema, abrindo a possibilidade de que um *link* DOI possa ser resolvido independentemente de

onde ele for encontrado, mesmo que ele apareça num recurso que explicitamente não suporte padrões de *open link*, ou num documento no formato PDF armazenado no computador do usuário. Isto se torna viável pela existência do *proxy DOI*, através do qual requisições de resolução podem ser redirecionadas para um componente de serviço, ou seja, para um resolvidor *OpenURL* no contexto de uma instituição. Vejamos o fluxo a seguir à guisa de exemplo (Beit-Arie et al., 2001):

1) Um leitor clica num *link DOI* encontrado na forma de citação no artigo de um periódico eletrônico. Este URL é direcionado ao *proxy DOI*, que é o alvo de todos os *links DOI* no contexto DOI/CrossRef, juntamente com um *cookie* previamente instalado no navegador do leitor, que expressa a componente de serviço - a *baseURL* - preferida por este leitor.

2) O *proxy* está habilitado a reconhecer o resolvidor local identificado pelo *cookie*. A partir desse dado, monta e transmite para este resolvidor local um *OpenURL* que transporta o DOI do artigo. Cabe assinalar que, caso a requisição venha de um usuário sem acesso à componente de serviço, o DOI será resolvido pelo processo *default* de resolução.

3) O resolvidor local pode então usar o DOI transmitido pelo *OpenURL* como argumento para consultar a base de dados CrossRef e recuperar os metadados do artigo.

4) Usando os metadados pode então formar uma "expressão" *OpenURL* e disponibilizar serviços estendidos, apropriados para o usuário que clicou o *link/DOI*, o qual, por sua vez, pode explorar resoluções alternativas para o DOI.

Pode-se inferir, a partir do fluxo acima, que os editores tornam seus produtos aderentes ao *OpenURL* quando usam o sistema CrossRef/DOI para identificar seus conteúdos. Uma implicação disso é a tendência dos editores acadêmicos a utilizarem o DOI como o principal mecanismo de *link* para texto completo (Brand, 2001; O'Neill, 2003).

CONCLUSÃO

O sucesso atual dos sistemas distribuídos de informação, tal como a World Wide Web, bem como o potencial de desenvolvimento futuro desses sistemas, depende fortemente da consistência e estabilidade dos *links* entre recursos digitais que eles disponibilizam *on-line*, via rede de computadores. A identificação persistente é uma característica chave para os repositórios digitais - sejam eles bibliotecas ou arquivos - que precisam funcionar segundo os paradigmas de interoperabilidade. Não se pode mais prescindir dessa característica quando se pensa na integridade dos objetos digitais e nas várias faces do acesso contínuo a eles. As formas de acesso devem contemplar tanto o acesso livre, o contexto do usuário, como também a proteção e a gestão dos direitos, com o estabelecimento de uma infra-estrutura para comércio eletrônico de conteúdos digitais. Uma arquitetura que inclua esquemas de nomes persistentes, somados a uma estrutura de metadados, é um imperativo para a viabilização de uma variedade de transações avançadas, as quais incluem busca distribuída, busca sensível ao contexto pessoal e institucional dos usuários, resolução múltipla e seleção e uso de objetos digitais. Felizmente, as soluções propostas e as já implementadas são convergentes e guardam entre si um alto grau de compatibilidade, posto que têm como base uma infra-estrutura comum, que é o URN. Mas é necessário dizer que as vinculações entre recursos digitais via *links* referenciais é contínua, sendo estes o elo mais fraco e mais instável, quando consideramos os fluxos de interoperabilidade entre sistemas e serviços de informação na Internet. No entanto, são eles que, garantindo o acesso contínuo e contribuindo para a preservação do acesso e da integridade dos conteúdos digitais, irão permitir a plena realização do conceito de bibliotecas digitais distribuídas, de arquitetura aberta e interoperáveis.

REFERÊNCIAS

ATKINS, H. et al. Reference linking with DOI's. *D-Lib Magazine*, v.6, n.2, 2000. Available from: <<http://dlib.org/dlib/february00/02risher.html>>. Cited: 3 Aug. 2006.

ARMS, W.Y; RESTON, V. Key concepts in the architecture of the digital library. *D-lib Magazine*, 1995. Available from: <<http://www.dlib.org/dlib/July95/07arms.html>>. Cited: 29 Aug. 2006.

- BEIT-ARIE, O. et al. Linking to the appropriate copy. *D-Lib Magazine*, v.7, n.9, 2001. Available from: <<http://www.dlib.org/dlib/september01/caplan/09caplan.html>>. Cited: 4 Jul. 2006.
- BERNERS-LEE, T.; FIELDING, R.; IRVINE, U.C.; MASINTER, L. *RFC2396 Uniform Resource Identifier (URI): generic syntax*. Network Working Group, 1988. Available from: <<http://rfc.net/rfc2396.html>>. Cited: 28 Aug. 2006.
- BIGWOOD, D. Persistent links, one solution to a common problem. *HAL-PC Magazine*, 1999. Available from: <<http://www.hal-pc.org/journal/june99/Columns/links/links.html>>. Cited: 29 Aug. 2006.
- BRAND, A. CrossRef TURN's one. *D-Lib Magazine*, v.7, n.5, 2001. Available from: <<http://www.dlib.org/dlib/may01/brand/05brand.html>>. Cited: 16 Jun. 2006.
- BRAND, A. CrossRef after three years. *ICSTI Forum*, n.44, 2003. Available from: <<http://www.icsti.org/forum/44/index.html#crossref>>. Cited: 2 Aug. 2006.
- CAPLAN, P.; ARMS, W.Y. Reference linking for journal articles. *D-Lib Magazine*, v.5, n.4, 1999. Available from: <<http://www.dlib.org/dlib/july99/caplan/07caplan.html>>. Cited: 4 Jul. 2006.
- CLEVELAND, G. *Digital libraries: definitions, issues and challenges*. IFLANET UDT Occasional Papers, 1998. Available from: <<http://www.ifla.org/VI/5/op/>>. Cited: 29 Aug. 2006.
- CROSSREF.ORG. *OpenURL and CrossRef*. Available from: <<http://www.crossref.org/02publishers/16openurl.html>>. Cited: 15 Jul. 2004.
- DACK, D. *Persistent identification systems, part 1: background*. National Library of Australia 2001. Available from: <<http://www.nla.gov.au/initiatives/persistence/Plpart1.html>>. Cited: 09 Aug. 2006.
- DOI system overview. International DOI Foundation, 2006. Disponível em: <http://www.doi.org/overview_sys_overview_021601.html>. Cited: 30 Aug. 2006.
- INTERNATIONAL DOI FOUNDATION. *DOI handbook*, 2006. Available from: <http://www.doi.org/handbook_2000/DOIHandbook-v4-3.pdf>. Cited: 30 Aug. 2006.
- KOEHLER, W. Web page change and persistence: a four-year longitudinal study. *Journal of the American Society for Information Science and Technology*, v.53, n.2, p.162-171, 2002. (DOI:10.1002/asi.10018).
- LANGSTON, M.; TYLER, J. Linking to journal articles in an online teaching environment: the persistent link, DOI, and OpenURL. *The Internet and High Education*, v.7, p.51-58, 2004.
- LIBRARY OF CONGRESS. *The relationship between URN's, Handles, and PURL's*. 1997. Disponível em <<http://memory.loc.gov/ammem/award/docs/PURL-handle.html>>. Cited: 10 Aug. 2006.
- LANNOM, L. Handle system overview. In: IFLA COUNCIL AND GENERAL CONFERENCE, 66., 2000, Jerusalem. *IFLA Conference Proceedings*. The Hague: Iplanet, 2000. p.1-6. Available from: <<http://www.ifla.org/IV/ifla66/papers//03282e.htm>>. Acesso em: 28 Sept. 2006.
- NATIONAL LIBRARY OF AUSTRALIA. *Managing web resources for persistent access* 2002. Available from: <<http://www.nla.gov.au/guidelines/persistence.html>>. Cited: 23/08/2006.
- MOATS, R. RFC 2141 URN syntax. Network Working Group, 1997. Available from: <<http://www.rfc-archive.org/getrfc.php?rfc=2141>>. Cited: 23 Aug. 2006.
- O'NEIL, J. Reference linking: A basic overview. *Quarterly Newsletter of the International Council for Scientific and Technical Information*, n. 44, 2003. Available from: <<http://www.icsti.org/forum/44/index.html#overview>>. Cited: 31 Aug. 2006.
- PASKIN, N. DOI: a 2003 progress report. *D-Lib Magazine*, v.9, n.6, 2003. Disponível em <<http://www.dlib.org/dlib/june03/paskin/06paskin.html>>. Acessado em 13 Jul. 2006.
- PAYETTE, S. Persistent identifiers on the digital terrain. *RLG Diginews*, v.2, n.2, 1998. Available from: <<http://www.rlg.org/preserv/diginews/diginews22.html#Identifiers>>. Cited: 9 Aug. 2006.
- PENTZ, Ed. A Collaborative linking network. *Issues in Science and Technology Librarianship*, n.29, 2001. Available from: <<http://www.library.ucsb.edu/istl/01-winter/article1.html>>. Cited: 16 May. 2006.
- PETKE, R.; KING, I. RFC 2717 Registration procedures for URL scheme names. Network Working Group, 1999. Available from: <<http://rfc.net/rfc2717.html>>. Cited: 29 Aug. 2006.
- POWELL, A. OpenResolver: a simple OpenURL resolver. *Ariadne*, n.28, 2001. Available from: <<http://www.ariadne.ac.uk/issue28/#resolver/>>. Cited: 12 Jul. 2006.
- SAYÃO, L. F. Preservação digital no contexto das bibliotecas digitais. In: MARCONDES, C.H.; KURAMOTO, H.; TOUTAIN, L.B.; SAYÃO, L.F. (Org.). *Bibliotecas digitais: saberes e práticas*. Salvador/Brasília: UFBA/IBICT, 2006. p.115-149.
- SHAFER, K. et al. *Introduction to persistent Uniform Resource Locators*. OCLC Online Computer Library Center, 1996. Available from: <http://www.isoc.org/inet96/proceedings/a4/a4_1.htm>. Cited: 29 Aug. 2006.
- SOLLINS, K; MASINTER, L. RFC 1737 Functional requirements for Uniform Resource Names. Network Working Group, 1994. Available from: <<http://www.ietf.org/rfc/rfc1737.txt>>. Cited: 10 Aug. 2006.
- SOLLINS, K. RFC 2276 Architectural principles of Uniform Resource Name Resolution. Network Working Group, 1988. Available from: <<http://rfc.net/rfc2276.html>>. Cited: 29 Aug. 2006.

VAN de SOMPEL, Herbert. Reference linking in a hybrid library environment: Part 1 - Framework for linking. *D-Lib Magazine*, v.5, n.4, 1999. disponível em <http://www.dlib.org/april99/van_de_sompel/04van_de_sompel-pt1.html>. Cited: 4 Jul. 2006.

VAN de SOMPEL, Herbert; HOCHSTENBACH, Patrick; BEIT-ARIE, Oren. *OpenURL syntax description*. 2000. Available from: <<http://www.sfxit.com/openurl/openurl.html>>. Cited: 30 Aug. 2006.

VAN de SOMPEL, H.; BEIT-ARIE, O. Open linking in the scholarly information environment using the Open URL framework. *D-Lib Magazine*, v.7, n.3, 2001. Available from: <<http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>>. Cited: 13 Jul. 2006.

WEIBEL, Stuart; JUL, Eric; SHAFER, Keith. *PURL's: Persistent Uniform Resource Locator*. OCLC Online Computer Library Center 1995. Available from: <http://purl.oclc.org/docs/new_purl_summary.html>. Acessado em 29 Aug. 2006.