

“LINKED DATA” – DADOS INTERLIGADOS - E INTEROPERABILIDADE ENTRE ARQUIVOS, BIBLIOTECAS E MUSEUS NA WEB

Carlos Henrique Marcondesⁱ

Resumo: Catálogos Web em sistemas de arquivos, bibliotecas e museus são hoje recursos informacionais fechados, que utilizam tecnologias, padrões e interfaces próprias, não permitindo navegar via diferentes recursos dentro dos catálogos e vice-versa. Tecnologias “Linked Data” – dados interligados –, parte da proposta da Web Semântica e oferecem a possibilidade de interligar recursos informacionais Web através de links semânticos, permitindo aos usuários uma navegação natural e intuitiva, seguindo esses links, por esses recursos, independentemente de interfaces de consulta específicas. O objetivo deste artigo é identificar e discutir as potencialidades oferecidas pelas tecnologias da Web Semântica, em especial, “Linked Open Data”, para que arquivos, bibliotecas e museus disponibilizem e tornem interoperáveis seus acervos na Web. Será utilizada metodologia de caráter qualitativo, do tipo “levantamento do estado da arte”, tendo como método a pesquisa bibliográfica, a visita a “sites” de interesse e a análise do material levantado.

Palavras-chave: Web Semântica. Dados interligados. Linked Open Data. Interoperabilidade. Padrões. Catálogos Web. Arquivos. Bibliotecas. Museus.

LINKED DATA AND INTEROPERABILITY BETWEEN ARCHIVES, LIBRARIES AND MUSEUMS ON THE WEB

Abstract: Nowadays, Web catalogs in archives, libraries and museums systems are locked information resources which use their own technologies, interfaces and standards, which do not allow the navigation to occur from outside to the catalog and vice-versa. Linked Data technologies, which are a component of Semantic Web proposal, offer the possibility of interlink information resources through semantic links, thus providing users with a natural and intuitive way to navigate, independently from specific systems or interfaces. The objective of this article is to identify the potentiality of the Semantic Web technologies, in special Linked Data, in order to provide interoperability between catalogs in archives, libraries and museums. We use a qualitative methodology to achieve a state-of-the-art on the issue, using bibliographic research methods and visiting sites to collect data.

Keywords: Semantic Web. Linked Open Data. Interoperability. Standards. Web catalogs. Archives. Libraries. Museums.



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/)

ⁱ Universidade Federal Fluminense. marcon@vm.uff.br.

Recebido em: 01/05/2012; aceito para publicação em: 17/07/2012.

1 INTRODUÇÃO

Todo ano é disputada, nas águas da baía de Guanabara, no Rio de Janeiro, a Regata da Escola Naval, uma instituição militar. Imagine que um jovem velejador australiano pretende disputar a regata; seu pai é um acadêmico da área de História, de origem portuguesa, com interesse em história colonial, e está abrindo uma nova linha de pesquisa na sua universidade, a saber, História Comparada entre países de colonização britânica e ibérica. Sua mãe, também acadêmica, é antropóloga com interesse em povos primitivos e na disseminação do *homo sapiens* por todos os continentes; ela tem origens suíças, sendo seus pais originários da cidade de Berna.

Graças a seu conhecimento da língua, o pai procurou no *site* do programa esportivo da TV Brasil – a televisão pública brasileira – dados sobre a regata, acomodações no Rio de Janeiro, passagens etc. Com sua curiosidade de acadêmico e a facilidade de conhecer a língua portuguesa, descobriu que a Escola Naval no Rio de Janeiro situava-se na ilha de Villegagnon; o *sítio Web* da TV Brasil trazia um *link* para o verbete da Wikipédia sobre Villegagnon, quando nosso curioso pai descobriu ser aquele um nobre francês, o cavaleiro de Malta, fundador da França Antártica, outro assunto relacionado ao verbete sobre Villegagnon (“skos:related”, SKOS, 2004). O verbete sobre a França Antártica esclarecia ser este um empreendimento colonial francês que pretendia estabelecer uma colônia exatamente na ilha de Villhegagnon, na baía de Guanabara. Essa empreitada era patrocinada pelo almirante francês Jean de Colliny, então um dos líderes do partido reformista na França, que acabou assassinado na célebre “Noite de São Bartolomeu”; Colliny pretendia estabelecer uma colônia que servisse de refúgio aos calvinistas, perseguidos na França da época.

Do verbete da Wikipédia em português, nosso curioso pai acessou, no catálogo da Biblioteca Nacional Brasileira, livros sobre temas relacionados à Villegagnon e à França Antártica. Ali baixou textos, pôde ler e se informar sobre a França Antártica, soube que os franceses haviam se aliado aos indígenas tupinambás, uma tribo que vivia na região, contra os portugueses; seguiu o *link* desse assunto desde o catálogo da Biblioteca Nacional e do cabeçalho de assunto “tupinambá”, e chegou ao assunto “tupinambá, tribo”, de mesmo significado (owl:sameAs, OWL, 2004), do catálogo de um tal Museu do Índio, que ele descobriu estar situado na mesma cidade do Rio de Janeiro, com importantes coleções de artefatos sobre a tribo dos Tupinambás, de grande interesse para os estudos da esposa. Descobriu também que vários personagens ligados ao projeto da França Antártica tiveram

relação com Calvino, o reformador religioso, residente na cidade de Berna, origem dos pais de sua esposa – pôde acessar o texto da correspondência trocada entre Calvino e Villegagnon (TAVARES, 2011).

Toda essa navegação foi orientada por *links* que lhe informavam o que encontrar, se fossem seguidos, permitindo assim uma navegação segura entre as diferentes fontes de informação que tinham, como curadores, diferentes instituições. Parece uma ficção? Realmente o é, mas essa é exatamente a proposta *Linked Open Data* – literalmente, interligar dados abertos –, que vem se tornando uma realidade ao acenar com a possibilidade de interligar acervos em arquivos, bibliotecas e museus digitais através de tecnologias da Web Semântica como *Resource Description Framework* (RDF) (RDF PRIMER, 2004) e *Uniform Resource Identifiers* (URIs) (RFC 2396, 1998).

A proposta de dados abertos interligados oferece grande potencial ao conectar recursos informacionais através de *links* semânticos, *links* que são significativos também para programas. Ao contrário, *links* convencionais nada mais são (além de uma eventual etiqueta textual significativa para usuários humanos) que meios para que programas navegadores, a partir de um recurso, acessem outro, *sem explicitar qual o significado da ligação entre os recursos*. Sendo significativos para programas, *links* semânticos podem ser processados de forma mais rica por eles, explorando e enriquecendo cognitivamente o significado (legível por máquina) da ligação entre ambos os recursos.

Assim poderíamos registrar, de forma significativa para um programa navegador (e este para seu usuário), que o significado do *link* que une a página da Wikipédia sobre Machado de Assis à página da minissérie da TV Globo *Capitu* é que Machado de Assis é o autor do livro *Dom Casmurro* que, por sua vez, tem uma versão em vídeo chamada *Capitu*. O significado do *link* entre o verbete de Machado de Assis na Wikipédia e a página da minissérie *Capitu* é óbvio para quem, como nós, vivemos no Brasil, lemos e fizemos fichas de leitura de *Dom Casmurro* desde os bancos escolares e assistimos à Rede Globo; mas não é, de forma nenhuma, óbvio para um internauta da Austrália, da China ou da Alemanha. A figura seguinte ilustra essa situação; vê-se como os *links* semânticos podem agregar valor à navegação de um internauta.

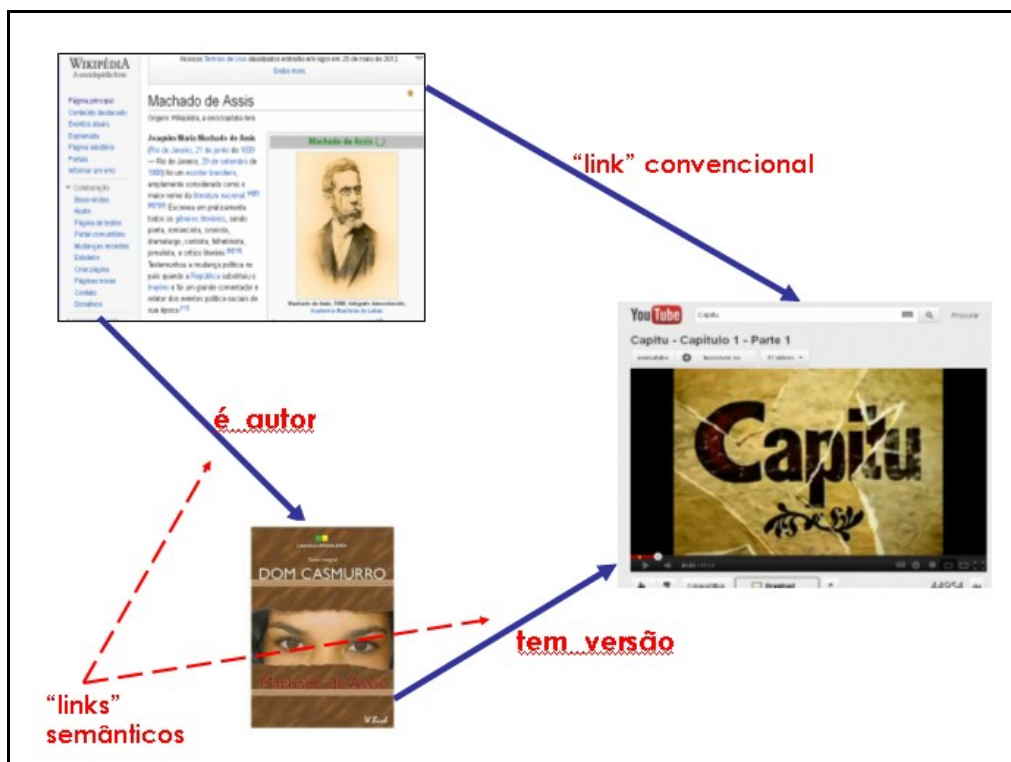


Figura 1: Links convencionais X links semânticos

Para viabilizar *links* semânticos, são utilizados padrões da Web Semântica, RDF e URI: “The glue that holds together the traditional document Web is the hypertext links between HTML pages. The glue of the data web is RDF links” (BIZER et al., 2007). Na prática, a proposta de dados abertos interligados significa, segundo Isele e colaboradores (2010): "First, to employ the RDF data model to publish structured data on the Web. Second, to [use http URIs] to set explicit RDF links between data items within different data sources". Assim, as tecnologias de dados interligados não são propriamente tecnologias de recuperação de informações, mas sim, uma vez recuperado um recurso informacional, tecnologias para interligá-lo com outros que lhe agreguem valor semântico, cultural, cognitivo, econômico ou científico. A dimensão atual da proposta de dados interligados, mostrando as instituições e seus recursos de informação já envolvidos é comentada por Bizer, Heath e Berners-Lee e pode ser vista em <http://linkeddata.org/>.

A *Web* vem oferecendo possibilidades extraordinárias para que arquivos, bibliotecas, centros de documentação e museus possam ampliar sua oferta de serviços e produtos a seus usuários. Grande parte dessas instituições vêm disponibilizando seus acervos através da *Web*. Tal situação coloca para essas entidades a questão de ampliarem e repensarem os serviços que prestam a seus usuários, mudando o foco, até agora centrado somente nos usuários que iam,

fisicamente, à instituição. Para arquivos, bibliotecas e museus, a disponibilização de seus acervos através da *Web* está se constituindo, cada vez mais, na forma corrente de prestação de seus serviços a uma gama muito mais ampla de usuários.

Essas possibilidades podem agora se ampliar significativamente através das tecnologias de dados abertos interligados. Após uma primeira fase em que arquivos, bibliotecas e museus disponibilizaram seus acervos na *Web*, como tem sido a adoção, por essas instituições, das possibilidades agora oferecidas pelas tecnologias de dados abertos interligados? Qual o potencial dessas tecnologias para interligar e tornar interoperáveis acervos em arquivos, bibliotecas e museus? Quais são os casos exemplares? Quais ferramentas e metodologias existem?

Este ensaio tem como objetivo apresentar as potencialidades da proposta da Web Semântica, em especial as tecnologias de dados abertos interligados, para a disponibilização e interoperabilidade entre conteúdos em arquivos, bibliotecas e museus através da *Web*. Para isso, a partir de uma palestra sobre o tema e de informações disponíveis no sítio do projeto de “Datos Enlazados”, da Biblioteca Nacional da Espanha², foram coletados documentos fundamentais sobre a tecnologia de dados interligados aplicados a bibliotecas, que conduziram a outros documentos e experiências que estão sendo desenvolvidas; analisou-se também documentação do W3C³, que descreve as tecnologias que são a base da proposta *Linked Open Data*, como RDF e URI. Foram analisadas também, de forma comparativa, metodologias e tecnologias específicas usadas em catálogos de arquivos, bibliotecas e museus, como MARC, Z39.50, EAD, ISAD(G), CCAA2, OAI-PMH, Dublin Core etc.

O texto do artigo está estruturado como se segue: após essa introdução, a seção 2 discute brevemente a proposta da Web Semântica, para situar aí as tecnologias de dados abertos interligados como proposta de *links* semânticos e levantar a questão da semântica dos catálogos em arquivos, bibliotecas e museus. A seção 3 analisa os padrões convencionais em que se baseiam catálogos digitais em arquivos, bibliotecas e museus. A seção 4 exemplifica como seria a conversão de um catálogo de biblioteca para a proposta de dados abertos interligados e discute suas potencialidades para integrar acervos em arquivos, bibliotecas e museus na Web. Finalmente, a seção 5 apresenta considerações finais e conclusões.

² <http://www.bne.es/es/Catalogos/DatosEnlazados/index.html>

³ W3C, World Wide Web Consortium, <http://www.w3.org>

2 WEB SEMÂNTICA E A PROPOSTA DE DADOS ABERTOS INTERLIGADOS

A proposta da Web Semântica descreve uma *Web* qualitativamente diferente da atual, na qual, ao invés dos conteúdos serem estruturados para serem exibidos e lidos por pessoas, terão uma semântica “inteligível” por programas, os chamados “agentes de *software*”; ou seja, programas que não sejam criados especificamente para processarem somente um tipo determinado de dado. As tecnologias da Web Semântica são a evolução da *Web* atual, em direção a mais amplas possibilidades de interação e de recuperação semântica da informação. A *Web* atual é chamada de Web Sintática pelos formuladores do projeto da Web Semântica por trabalhar apenas com padrões sintáticos de *forma*. Uma busca por conteúdo é processada por programas como uma busca de padrões de textos *idênticos* aos fornecidos como parâmetro, que são, em última instância, meros padrões de *bits* acesos ou apagados. Essa mudança de qualidade é destacada por Tim Berners-Lee em vários de seus textos:

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. (BERNERS-LEE, 2001 p. 2).

E ainda:

The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users (BERNERS-LEE, 2001 p. 2).

Longe de se constituir somente numa proposta meramente tecnológica, a proposta da Web Semântica é bastante ampla e ambiciosa:

The Semantic Web is not “merely” the tool for conducting individual tasks that we have discussed so far. In addition, if properly designed, The Semantic Web can assist the evolution of human knowledge as a whole (BERNERS-LEE, 2001 p. 2).

Mas o que significaria, exatamente, uma “semântica” inteligível por programas? As diferentes linguagens-padrões da Web Semântica para estruturação de conteúdos são apresentadas em (OWL, 2004) por ordem crescente de “expressividade semântica”, ou seja, dispõe de um vocabulário capaz de expressar um leque cada vez mais amplo de significados:

- XML - sintaxe para estruturar documentos. Não impõe restrições semânticas.
- XML Schema - restringe a estrutura de documentos XML
- RDF - modelo de dados para objetos e relacionamentos. Semântica simplificada
- RDF Schema - vocabulário para descrever propriedades e classes expressas em RDF; semântica para generalização.
- OWL - maior vocabulário para descrever classes e propriedades: relacionamentos (p. ex: disjunção), cardinalidade (p. ex: exatamente um), igualdade, tipos mais

expressivos de propriedades, características de propriedades (p. ex: simetria) e enumeração dos elementos de uma classe⁴.

Segundo Berners-Lee (2001), o padrão RDF seria fundamental e marcaria um ponto de inflexão na sequência de linguagens apresentada, pois ali estaria o primeiro “lampejo” do que seria uma “semântica” inteligível por programas. “In short, XML allows users to add arbitrary structure to their documents but says nothing about what the structures mean. **Meaning is expressed by RDF**⁵, which encodes it in sets of triples, each triple being rather like the subject, verb and object of an elementary sentence”. O RDF é justamente o padrão básico da proposta de dados abertos interligados, através do qual recursos são descritos, por meio de uma ou mais assertivas sobre eles; conteúdos são publicados de forma estruturada, enquanto uns e outros são interconectados através de *links* diferentes dos convencionais, *links* que expressam a natureza ou “semântica” da ligação entre os recursos sendo ligados (BIZER et al., 2007). A “semântica” inteligível por programas se daria porque o RDF seria, no ordenamento das linguagens da *Web* mencionado, a primeira linguagem (a segunda seria a OWL) que pressupõe uma semântica; esta é composta por sujeito, predicado e objeto de uma assertiva, os elementos de seu metamodelo. Este é definido, nas palavras de Guizzardi (2005, p. 36) da maneira que segue:

What is referred by structure of a language can be accessed via the description of the specification of conceptual model underlying the language, i.e., a description of the worldview embedded in the language’s modeling primitives. In (Milton & Kamierczak, 2004), this is called the ontological metamodel of the language, or simply, the ontology of the language. From a philosophical standpoint, this view is strongly associated with Quine (Quine, 1969), who proposes that an ontology can be found in the ontological commitment of a given language, that is, the entities the primitives of a language commit to the existence of. For example, Peter Chen’s Entity Relationship model (Chen, 1976) commits to a worldview that accounts for the existence of three types of things: entity, relationship and attribute.

Assim, um metamodelo reúne os pressupostos semânticos – o significado – de cada elemento do vocabulário da linguagem. Segundo o metamodelo ontológico da linguagem RDF (ou seja, que a semântica pode ser expressa com os elementos da linguagem RDF), sempre se poderá identificar, numa assertiva sobre qual entidade é feita a assertiva (o sujeito), que tipo de assertiva está sendo dita sobre essa entidade (o predicado) e o que, especificamente, está sendo dito sobre a entidade (o objeto) (RDF PRIMER, 2004). A sentença “O autor da página <http://www.professores.uff.br/marcondes/publicacoes.htm>” é Carlos Marcondes” seria expressa em RDF da seguinte maneira:

⁴ Tradução nossa.

⁵ Grifo nosso.

```

<?xml version="1.0" >
<rdf:RDF
  xmlns= http://www.w3.org/1999/02/22-rdf-syntax-ns
  xmlns:dc="http://purl.org/dc/elements/1.1">
  <rdf:Description rdf:about
    "http://www.professores.uff.br/marcondes/publicacoes.htm">
    <dc:creator>MARCONDES, Carlos</dc:creator>
    <dc:date>2005</dc:date>
  </rdf:Description>
</rdf:RDF>

```

Figura 2. Expressão RDF para “Carlos Marcondes é o autor da página <http://www.professores.uff.br/marcondes/publicacoes.htm>”

No exemplo, a assertiva em RDF utiliza também o que é chamado de *name spaces*, espaços de nomes, que seriam vocabulários específicos para designar o predicado e o objeto da assertiva. Os predicados *autor* e *data de publicação* são expressos usando os elementos do vocabulário de metadados descritivos Dublin Core, “dc:creator” e “dc:date”; este vocabulário é identificado univocamente dentro do documento RDF através da declaração do espaço de nomes Dublin Core, “xmlns:dc=“http://purl.org/dc/elements/1.1”. O recurso *espaço de nomes* permite que elementos de mesmo nome possam ser utilizados com outro significado dentro do mesmo documento RDF, desde que pertençam a *outro* espaço de nomes diferente, como, por exemplo, o autor de um livro junto ao autor de uma ação judicial.

Ao contrário de semânticas locais, embutidas e prisioneiras em programas específicos, a proposta da Web Semântica faz sua aposta em linguagens que estruturam os conteúdos de forma crescente, provendo relações semânticas cada vez mais complexas, que permitam a programas *seguir* essas relações realizando “inferências computacionais”. Com essas linguagens crescentemente expressivas, pode-se modelar/representar, no espaço computacional, domínios complexos – as chamadas ontologias (LEGG, 2007) – e processar essas representações por programas. Esses conteúdos estruturados segundo essas linguagens

padronizadas devem estar o “mais público possível”, disponíveis na própria *Web*, permitindo que programas genéricos – os agentes inteligentes (THE SEMANTIC WEB, 2003) – os processem. Esses programas genéricos têm um mínimo de semântica embutida (ou seja, prisioneira) em seus códigos, somente o necessário para processarem inferências – *seguirem* as relações nos conteúdos estruturados nas linguagens padrão da Web Semântica.

Nas assertivas em RDF, tanto o recurso (o sujeito da assertiva) quanto o predicado e o objeto podem ser identificados univocamente em todo o espaço da *Web* por URIs (RFC 2396, 1998).

Ao contrário dos URLs, que são endereços de recursos como, por exemplo, de uma página hospedada num servidor (<http://www.professores.uff.br/marcondes/publicacoes.htm>), os URIs são identificadores únicos no espaço de um dado servidor, resultando, portanto, em identificadores únicos no espaço da *Web*.

Quando utilizados para conectar recursos, os URIs resultam em *links* permanentes e estáveis, além de terem um significado preciso, ao contrário dos URLs, meros endereços de páginas, instáveis e sempre sujeitos ao frustrante “Erro 404 Page not found”, caso a organização interna de pastas no servidor que abriga o recurso mude por algum motivo ou mude o domínio do servidor. Ao invés dos *links* para outras páginas como na *Web* convencional, na proposta *Linked Open Data* os *links* são para *recursos* como, por exemplo, o registro de um livro num catálogo de biblioteca, ou um verbete numa enciclopédia, e não para meras páginas hipertextuais; além disso os *links* têm *significado*, são *links* semânticos, expressam a relação entre os recursos linkados.

Os URIs podem identificar dois tipos de recursos: os recursos que têm existência na *Web*, como um documento digital, uma imagem digital ou uma base de dados, os chamados recursos informacionais; e os recursos que não têm existência na *Web*, como eu, você, uma empresa, uma peça num acervo museológico, um livro raro numa biblioteca, ou seja, os chamados recursos não informacionais. No último caso, o recurso propriamente dito deve ser *substituído* por uma representação dele mesmo. Essa distinção é fundamental no ambiente de dados interligados porque os dois tipos de recursos têm tratamentos distintos quando acessados: ao receber de um navegador uma requisição para acessar um recurso informacional, o servidor onde o recurso está hospedado responde enviando uma cópia digital deste. Diferentemente, uma vez que recursos não informacionais não podem ser acessados diretamente, ao receber de um navegador uma requisição para acessar um recurso não informacional, o servidor onde as representações desse recurso estão hospedadas responde

informando que tipos de representações estão disponíveis, geralmente representações em RDF – legíveis por programas, portanto –, ou representações em HTML, legíveis por pessoas. O navegador escolhe então a representação mais adequada e esta lhe é enviada pelo servidor, ou seja, há uma negociação de quais conteúdos acessar entre os programas navegador e servidor.

Assim, um servidor de recursos preparado para a proposta de dados abertos interligados atende a três URIs para cada recurso não informacional: uma que identifica a cidade do Rio de Janeiro propriamente dita (um recurso não informacional), por exemplo <http://www.cidadesdobrasil.org.br/recursos/riodejaneiro>; outra, <http://www.cidadesdobrasil.org.br/dados/riodejaneiro>, para representações da cidade do Rio de Janeiro em RDF, e uma terceira, <http://www.cidadesdobrasil.org.br/paginas/riodejaneiro>, para uma página em HTML com informações sobre a cidade.

Diversos servidores podem manter representações distintas do mesmo recurso, seja ela uma entidade real ou informacional. Isso enriquece a *Web*, uma vez que pontos de vista e opiniões distintas podem ser expressas sobre a mesma entidade, coexistindo sob a forma de *links* e assertivas distintas sobre esta; *links* distintos sobre a mesma entidade podem ser combinados através do comando da linguagem OWL “owl:sameAs”, incluído nas descrições RDF do recurso.

Todo esse esquema oferece, assim, novas possibilidades de navegação qualificada, utilizando uma nova geração de programas navegadores semânticos, como o DISCO⁶ ou o OpenLink RDF Browser⁷, capazes de interpretar e seguir *links* semânticos. Isso permite a aquisição/descoberta e relacionamento entre diferentes recursos de conhecimento disponíveis na *Web* de forma cognitivamente mais rica.

3 SISTEMAS DE GERENCIAMENTO DE ACERVOS EM ARQUIVOS, BIBLIOTECAS E MUSEUS, INTEROPERABILIDADE E A QUESTÃO SEMÂNTICA

Sistemas e metodologias específicos utilizados por arquivos, bibliotecas e museus tornam seus catálogos fechados em si mesmos, com semânticas locais, dificultando a integração de seus conteúdos com outros arquivos, bibliotecas e museus, ou mesmo com outros recursos informacionais também disponíveis na *Web*. Apesar de estarem disponíveis na

⁶ DISCO Hyperdata Browser,

⁷ OpenLink RDF Browser,

Web, não têm praticamente *links* para outros recursos, não permitindo que se navegue através de *links* do catálogo para recursos externos e destes de volta para o catálogo. Durante anos essas instituições têm desenvolvido semânticas, modelos, formatos e padrões próprios. Todo o potencial de integração e acesso proporcionado pela *Web* não pode ser aproveitado pelos sistemas convencionais que gerenciam esses catálogos, impedindo que tais acervos possam aproveitar suas sinergias e potenciais culturais, e que se beneficiem, e também que outras instituições possam fazer *links* para seus acervos.

Interoperabilidade é definida por Miller (2000) como: “the ability of a system or a product to work with other systems or products without special effort on the part of the customer”. No caso que interessa a esse trabalho, a interoperabilidade entre catálogos de arquivos, bibliotecas e museus seria a capacidade desses sistemas de integrar seus conteúdos. As soluções de interoperabilidade existentes entre sistemas de arquivos, bibliotecas e museus entre si são pouco satisfatórias e complicadas tecnologicamente. O protocolo Z39.50⁸, usado para permitir consultas simultâneas a catálogos de diferentes bibliotecas, é um sistema que consome muitos recursos computacionais, exige padrões específicos da área biblioteconômica, como o MARC. A sincronização das consultas se dá na velocidade do servidor Z39.50, mais lento; esses padrões e as tecnologias a ele associadas remontam à década de 1990. O OAI-PMH⁹, protocolo de coleta automática de metadados de repositórios e bibliotecas digitais para depositá-los numa base de metadados centralizada, na qual se dá a consulta simultânea, embora seja uma solução tecnologicamente simples, reduz a descrição de documentos utilizada nos diferentes repositórios e bibliotecas digitais ao mínimo múltiplo comum do conjunto de elementos de metadados Dublin Core (2003), muito simples para aplicações mais sofisticadas.

Em se falando de interoperabilidade de sistemas de arquivos, bibliotecas e museus uns com os outros, a questão se torna ainda mais complexa. Enquanto objetos em bibliotecas e museus são individuais – o documento, a peça museológica –, os arquivos têm seu foco em *conjuntos* de documentos, os fundos. Mais complicado ainda é superar o isolamento dos sistemas de arquivos, bibliotecas e museus e integrá-los a outros sistemas existentes na *Web*, como enciclopédias, dicionários, bases de dados factuais, científicas e estatísticas, sistemas de georeferenciamento como o Geonames, bancos de imagens, de vídeos etc. Toda essa situação é resumida por Barmes (2011) da seguinte maneira: “Cultural heritage institutions maintain

⁸ Z39.50 Maintenance Agency, <http://www.loc.gov/z3950/agency/>

⁹ Open Archives Initiative, <http://www.openarchives.org/pmh/>

catalogues that exist like silos, isolated from one another, and isolated from the wide ecosystem of the web”.

Os sistemas de gerenciamento de conteúdos em arquivos, bibliotecas e museus foram desenvolvidos a partir das décadas de 1970 e 1980, antes do surgimento da *Web*. Os primeiros a serem criados, ainda na década de 1970, foram os *Integrated Library Systems* (ILSs), sistemas integrados de gestão de bibliotecas, que passam, na década de 1980, a disponibilizar seus *On-line Public Access Catalogs* (OPACs). Esses sistemas deram origem a um segmento forte da indústria de *software*. A primeira tentativa de produzir um sistema genérico, ou seja, não desenvolvido para atender às necessidades específicas de uma instituição de gestão de conteúdos para instituições arquivísticas, foi o projeto ICA-AtoM¹⁰, do Conselho Internacional de Arquivos, iniciado já no ano 2005. Na área de museus, apesar do avanço que significou a proposta do modelo conceitual CIDOC-CRM¹¹, ainda não existe consenso ou padrões que permitam o desenvolvimento de um pacote genérico de gestão de conteúdos museológicos.

Os conteúdos digitais gerenciados por sistemas de gestão de conteúdos de arquivos, bibliotecas e museus seguem padrões e formatos bastante específicos, que os tornam pouco interoperáveis e pouco suscetíveis de serem integrados a outros conteúdos diferentes, tais como música, vídeos, fotos etc. Ao acessar esses catálogos na *Web*, o usuário fica como que prisioneiro desses contextos sistêmicos e institucionais específicos, praticamente sem possibilidades de navegar de fora para dentro ou de dentro para fora destes. O formato MARC¹², no qual acervos gigantescos de milhares de bibliotecas pelo mundo estão registrados, está estruturado em identificadores numéricos específicos para cada campo, associados ao conteúdo propriamente dito do campo. Essa estrutura, o padrão ISO 2709, remota à década de 1980. Esse esquema reflete, antes de tudo, as necessidades de compartilhamento de recursos e gestão desses acervos pelas bibliotecas, e menos a preocupação com as necessidades de acesso amigável pelos usuários finais. Além dessas questões, do ponto de vista conceitual e ontológico, o formato MARC é confuso e carrega um grande grau de informalidade, na medida em que mistura, num único registro digital, conteúdos que pertencem a entidades distintas, como dados de autores, da própria obra, de editores, assuntos etc. Quando o MARC bibliográfico tradicional passa a ser utilizado fora do

¹⁰ ICA-AtoM, <http://ica-atom.org/>.

¹¹ The CIDOC Conceptual Reference Model, <http://www.cidoc-crm.org/>

¹² MARC Standard, <http://www.loc.gov/marc/>.

seu contexto original, para registrar autoridades ou para representar informação arquivística, o elenco de campos original é tomado para representar os novos conteúdos por analogias com estes; toda essa informalidade é compensada pelo conhecimento tácito de arquivistas e bibliotecários catalogadores, mas é desastrosa caso se pretenda que tais conteúdos possam ser processados de forma semanticamente consistente por programas.

Tal fato começa a ser minorado pela utilização de modelos ou ontologias como o FRBR (1998), que faz uma distinção clara entre obra, manifestação, expressão e item, o que não acontece no formato MARC, no qual atributos (campos de dados) dessas entidades estão misturados indistintamente. A diferenciação é essencial quando se pretende tratar esses conteúdos de forma generalizada, como na proposta da Web Semântica, para além das semânticas locais, embutidas e prisioneiras no código de programas específicos. Apesar da incorporação do FRBR pelo RDA (2010) – o novo código de catalogação que substituirá o CCAA2 –, a adoção do FRBR ainda é lenta, dado o legado de milhões de registros representados em formato MARC e segundo as regras do CCAA2.

A estrutura hierárquica de fundos, séries e dossiês, avalizada pela norma ISAD(G) (CONSELHO INTERNACIONAL DE ARQUIVOS, 2000), está baseada no princípio da proveniência, caro à gestão arquivística; acervos arquivísticos são mantidos obedecendo a esse princípio. Mesmo quando codificada em meio legível por computador usando XML¹³ (linguagem padrão do W3C para estruturar documentos de forma independente de programas específicos), o *Enconding Archival Description*¹⁴ (EAD) reflete essa estrutura hierárquica, opaca e pouco significativa para os usuários finais. Ambos os padrões refletem, antes, uma visão patrimonialista, de curadoria e gestão dos seus acervos, mais do que uma visão voltada para a disponibilização de conteúdos ao usuário final. A norma ISAD(G) também incorpora um grande grau de informalidade; entre outros casos, o campo para descrição multinível, criado para expressar a estrutura hierárquica de fundos, série e dossiês, não tem regras genéricas de preenchimento, sendo preenchido segundo convenções e semânticas locais das diferentes instituições, o que é um forte obstáculo para seu tratamento por máquina de forma generalizada, independentemente dos sistemas específicos de cada instituição.

Com a atual tecnologia em catálogos de arquivos, bibliotecas e museus, é possível tão somente fazer um *link* de um registro do catálogo para um recurso *Web* externo; é impossível fazer o *link* inverso, ou então o *link* de um registro do catálogo de uma biblioteca para o de outra, ou do registro do catálogo de uma biblioteca para o de um arquivo ou o de um museu.

¹³ Extended Markup Language, <http://www.w3.org/XML/>

¹⁴ EAD, *Enconding Archival Description*, <http://www.loc.gov/ead/>.

Tal situação se torna um empecilho à integração em larga escala do conteúdo de catálogos de arquivos, bibliotecas e museus, com a infinidade de outros recursos informacionais agora disponíveis na *Web*.

4 A PROPOSTA DE DADOS INTERLIGADOS E A INTEROPERABILIDADE DE ACERVOS EM ARQUIVOS, BIBLIOTECAS E MUSEUS NA WEB

Um projeto de dados interligados em uma biblioteca significa converter seu catálogo de MARC, por exemplo, para RDF. O metamodelo do RDF deve ser desdobrado num vocabulário ou ontologia de classes, a fim de identificar entidades de interesse e propriedades relacionando essas classes. Assim, sujeito e objeto devem ser desdobrados em classes e predicado em propriedades. Esse vocabulário pode (e geralmente é) composto por vários *name spaces* usados simultaneamente; a conversão do catálogo deverá identificar claramente o mapeamento entre os campos MARC e os elementos do vocabulário utilizado. Este, juntamente com os registros em formato MARC do catálogo, servem de insumo para programas que, com base nos conteúdos do registro MARC e na semântica proporcionada pelos elementos do vocabulário, convertem os registros do catálogo em triplas RDF.

No quadro a seguir, um registro MARC, referente ao Dom Casmurro de Machado de Assis, terá seus campos mapeados, num primeiro passo, para um vocabulário composto pelos elementos Dublin Core¹⁵ e, depois, na conversão final para RDF, para elementos tomados de outros vocabulários ou ontologias, como “owl:sameAs”¹⁶, “dcterms:hasVersion”¹⁷ e “edm:happenedAt”¹⁸:

Campo MARC	Campo Dublin Core
035_9(DLC)76076436	<dc:identifier> (DLC)76076436</dc:identifier>
1002_aMachado de Assis	<dc:creator>Machado de Assis</dc:creator>
245__aDom Casmurro	<dc:title>Dom Casmurro</dc:title>
260__aSão Paulo	

¹⁵ Name space dc: <http://purl.org/dc/elements/1.1/>

¹⁶ Name space owl: <http://www.w3.org/2002/07/owl#>

¹⁷ Name space dcterms: <http://purl.org/dc/terms/>

¹⁸ Name space edm: <http://www.europeana.eu/schemas/edm/>

260__ bEdições Melhoramentos	<dc:publisher>Edições Melhoramentos</dc:publisher>
260__ c1966	<dc:date>1966</dc:date>
651__ aRio de Janeiro	<dc:coverage>Rio de Janeiro</dc:coverage>

Quadro 1. Mapeamento entre campos MARC e elementos Dublin Core

Saber que o dc:creator de Dom Casmurro é alguém chamado “Machado de Assis” é pouco interessante para estrangeiros não versados em literatura brasileira; trata-se de um simples nome, sem maior significação. Muito mais interessante que isso para um internauta é navegar de “Dom Casmurro”, através de um *link* semântico cujo *significado* é dc:creator (dado pelo vocabulário Dubli Core), para um registro de um arquivo de autoridades como o VIAF¹⁹, no qual existe informação biográfica muito mais aprofundada, cuidadosamente curada, com menção às fontes, sobre Machado de Assis. Isso pode ser feito substituindo na assertiva RDF <dc:creator>Machado de Assis</dc:creator> pela URI do registro VIAF sobre Machado de Assis: <dc:creator>http://viaf.org/viaf/95151633/#Machado_de_Assis</dc:creator>.

O processo de conversão do catálogo MARC para RDF pode, de forma automática ou semi-automática, agregar valor aos registros bibliográficos, associando a estes todo um conjunto de *links* semânticos. No exemplo, o campo MARC 651 – assunto geográfico –, com conteúdo “Rio de Janeiro”, teria como mapeamento para o vocabulário Dublin Core <dc:coverage>Rio de Janeiro</dc:coverage>; esse campo Dublin Core pode ser reconhecido e convertido no *link* semântico do vocabulário edm (DEFINITION OF THE EUROPEANA DATA MODEL ELEMENTS, 2012) “happenedAt”, que conduziria ao mapa da cidade do Rio de Janeiro no sítio “geonames”²⁰. Seguindo esse *link* semântico, um internauta saberia que “Dom Casmurro” é ambientado na cidade do Rio de Janeiro; da mesma forma, o registro de Machado de Assis no VIAF poderia ser conectado através do “*link* semântico” “owl:sameAs” ao registro de Machado de Assis na Wikipedia/DBPedia (AUER, 2007); o registro bibliográfico poderia ser também conectado pelo *link* semântico do vocabulário dcterms “hasVersion” a capítulos da minissérie *Capitu*, produzida pela TV Globo, no YouTube²¹. Inúmeros outros *links* interessantes como esses podem ser estabelecidos, tornando

¹⁹ VIAF – Virtual International Authority File, <http://viaf.org>

²⁰ Geonames, <http://www.geonames.org/>

²¹ YouTube, <http://www.youtube.com>

consideravelmente mais rica a experiência de navegação pelos registros de catálogos de bibliotecas.

Toda essa situação pode ser ilustrada no grafo RDF da figura seguinte:

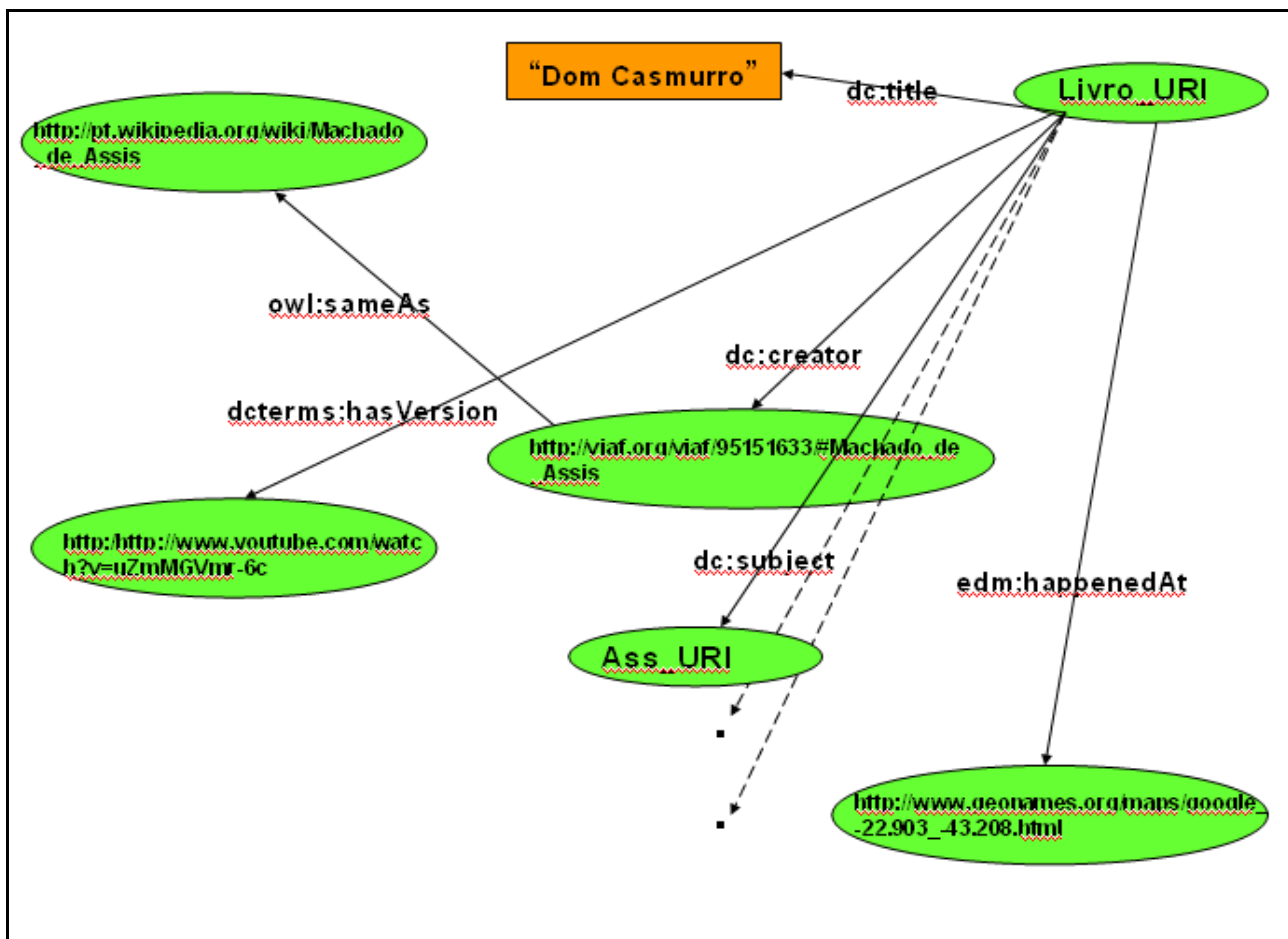


Figura 3. Grafo RDF ilustrando o exemplo descrito

Para que nosso exemplo fique mais próximo do real, vamos tomar o que chamamos de “Livro_URI” na figura, que seria a URI do livro Dom Casmurro de Machado de Assis em nossa hipotética biblioteca. O registro MARC apresenta o campo 035 – número de registro no sistema –, com o valor “(DLC)76076436”. Ora, um livro físico não possui um URI, ele é um objeto; teria que haver uma representação digital deste. Numa biblioteca preparada para a proposta de dados abertos interligados como a Biblioteca do Congresso dos EUA, essa representação existe; é o chamado “Permalink”, um URI para a ficha do livro Dom Casmurro, sua representação no espaço digital. Repare que essa ficha é identificada por um URI, independente portanto do sistema gerenciador do catálogo da LC; o “Permalink” do Dom

Casmurro é <http://lccn.loc.gov/76076436>. Se “colarmos” o “Permalink” da ficha do Dom Casmurro na janela de um navegador, a ficha será acessada independentemente de estarmos dentro do catálogo da LC, ou seja, cada ficha do catálogo da LC possui um URI, podendo ser linkada e acessada diretamente, independentemente do sistema de catálogo da LC.

A figura a seguir, retirada do sítio do projeto de “Datos Enlazados” da Biblioteca Nacional da Espanha²², dá uma visão geral dos vários passos metodológicos para a conversão do catálogo de uma biblioteca para essa tecnologia.

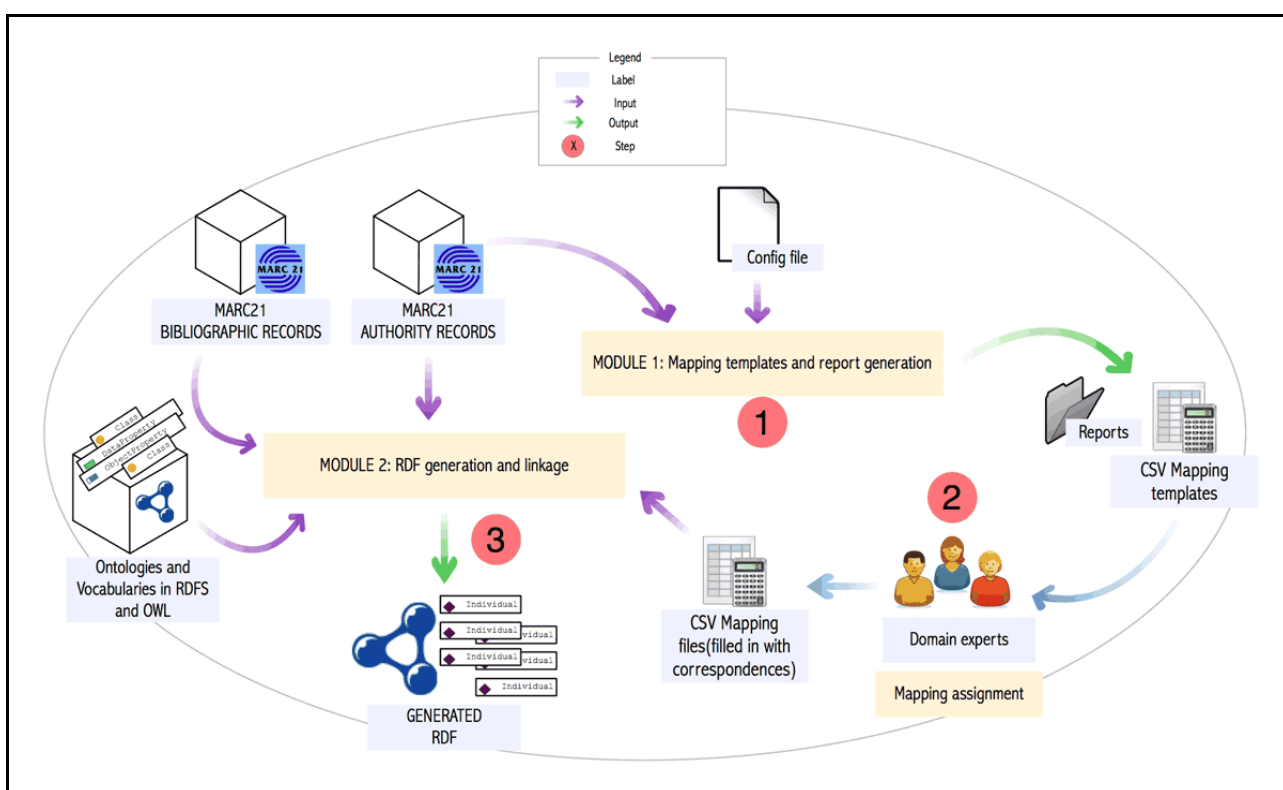


Figura 4. Passos metodológicos de um projeto de dados interligados

5 CONSIDERAÇÕES FINAIS

Num ambiente global como o de dados abertos interligados, a estabilidade/confiabilidade desses *links* é um fator essencial. Segundo o modelo proposto por Hannemann (2010), arquivos, bibliotecas e museus são entidades que tradicionalmente mantêm registros confiáveis sobre bens culturais sob sua custódia e são, portanto,

²² A figura foi obtida em http://www.bne.es/opencms/imagenes/CATALOGOS/Conversion_gr.gif.

vocacionadas para manter identificadores permanentes, URIs confiáveis e estáveis sobre esses identificadores. Além disso, nas suas tarefas como curadoras dos acervos de bens culturais, arquivos, bibliotecas e museus mantêm sistemas classificatórios, de autoridades, e terminológicos, que podem atuar como nós centrais para viabilizar a navegação por assuntos ou nomes que são comuns (BERMES, 2011) entre dados de catálogos de arquivos, bibliotecas e museus num ambiente de dados abertos interligados.

Assim, o *Library Linked Data Incubator Group Final Report* (2011) distingue, como dados de interesse a serem interligados em bibliotecas: *datasets*, ou seja, conjunto de dados bibliográficos ou o que seria o catálogo de uma biblioteca com sua base de autoridades; *element sets*, um vocabulário de metadados através do qual *datasets* seriam descritos; e *value vocabularies*, o conteúdo de uma base de autoridades. Estes últimos teriam um papel fundamental como nós centrais, interligando diferentes *datasets*, como foi dito anteriormente.

Outra questão facilitadora da interoperabilidade entre dados de arquivos, bibliotecas e museus é o alinhamento semântico (*Semantic alligment*). Isso significa ligar entidades equivalentes, similares ou semanticamente relacionadas (Library Linked Data Incubator Group Final Report, 2011) em bases de autoridades, vocabulários de metadados ou conjuntos de dados de diferentes instituições, como, por exemplo, o mesmo assunto ou nome em diferentes bases de autoridades. Alinhamentos semânticos facilitam a navegação através de dados abertos interligados entre recursos informacionais diferentes através dos *links* semânticos entre entidades em diferentes conjuntos de dados. Na área de arquivos, bibliotecas e museus, alguns esforços de alinhamento semântico já foram feitos, como entre os modelos FRBR (1998) (da área de bibliotecas) e o CIDOC CRM (da área de museus), resultando disso o FRBRoo (RIVA et al., 2008).

Várias bibliotecas pelo mundo estão implementando projetos de dados abertos interligados, entre elas a Biblioteca do Congresso dos EUA, a *British Library*, a Biblioteca Nacional da Alemanha, a Biblioteca do Centro George Pompidou na França, a Biblioteca Nacional da Espanha²³. No entanto, o projeto mais ambicioso de interligação de instituições arquivísticas, bibliotecas e museus é o da Biblioteca Europeia (Europeana Data Model Primer, 2010). Esse projeto se baseia na aquisição de metadados de objetos culturais de diferentes instituições arquivísticas, bibliotecas e museus através de mecanismos do protocolo OAI-PMH, bem como no enriquecimento desses metadados através das tecnologias de dados

²³ Maiores informações e “links” podem ser encontrados em <http://www.bne.es/es/Actividades/ActosCulturales/CicloCitaBN/CitaBN2011/dicdatosenzados.html>.

abertos interligados. Para permitir qualificar e enriquecer semanticamente tanto objetos culturais e/ou suas representações digitais quanto suas diferentes relações, o projeto Europeana desenvolveu um vocabulário específico – uma ontologia – de classes de objetos e processos e de relações entre estes, de interesse para instituições que mantêm acervos em cultura (Definition of The Europeana Data Model Elements, 2012). Os grafos RDF da Europeana fazem uso desse vocabulário.

Outro recurso importantíssimo para a iniciativa de dados abertos interligados é a versão em RDF da Wikipédia, a DBPedia (AUER, 2007). Vários recursos em Memória e Cultura que participam da iniciativa de dados abertos interligados têm *links* semânticos de e para verbetes da DBPedia. Pela sua natureza enquanto enciclopédia, a DBPedia funciona assim como um verdadeiro elo, ou ponte, interligando vários outros recursos específicos.

Um projeto de dados interligados demanda um leque de competências especiais. Nas palavras de Bowen (2010, p.):

Converting legacy metadata to linked data will require a team of experts, including MARC-based catalogers, specialists in other metadata schemas, software developers, and Semantic Web experts to design and test normalization/conversion algorithms, develop new schemas, and prepare individual records for automated conversion.

Ao se integrarem na *Web* através de “Linked Open Data”, arquivos, bibliotecas e museus aumentam a exposição de seus conteúdos, os quais podem passar a ser linkados desde fora de domínios temáticos específicos (por exemplo, documentos em arquivos, bibliotecas e museus sobre “história do Brasil”) desde fora do domínio tópico das instituições de cultura (como documentos de “história do Brasil” linkados a informações georreferenciadas sobre locais dos acontecimentos históricos, linkados com trechos de filmes sobre estes temas, linkadas, por sua vez, com sítios turísticos, agências de viagens e companhias aéreas, e com festivais de música que ocorrerão nesses locais). Ou seja, dados de arquivos, bibliotecas e museus passam a ser expostos e podem ser acessados por uma ampla e inusitada gama de novas audiências. Assim, segundo o Library Linked Data Incubator Group Final Report (2011), “all statements provided about a particular uniquely identified resource can be aggregated into a global graph”.

A navegação de um recurso ao outro através de *hiperlinks* é a forma natural para usuários descobrirem, acessarem e navegarem pela *Web*. Ao contrário, o acesso aos dados em catálogos de arquivos, bibliotecas e museus, com consulta a vocabulários controlados, uso de conectivos booleanos etc., é hoje a mesma que nos anos 1990, ou seja, é específica, totalmente estanque e isolada dos outros recursos existentes na *Web*. A proposta de dados

abertos interligados quando aplicada a arquivos, bibliotecas e museus oferece a possibilidade de que a navegação por dados em catálogos de tais instituições possa acontecer de forma mais natural para os usuários da *Web*. Através dela, arquivos, bibliotecas e museus podem integrar seus conteúdos de forma simples ao infossistema da *Web*, viabilizando a navegação “semântica” de um recurso a outro, rompendo assim seu isolamento e permitindo que consolidem seus papéis sociais.

Agradecimentos

Este artigo apresenta resultados parciais de um estágio pós-doutoral, desenvolvido em 2011, na Universidade Carlos III de Madrid, sob orientação do Prof. Dr. Jose Antonio Moreiro Gonzalez, com apoio da CAPES. Agradecemos também a Linair Maria Campos e Denise Lopes pelas valiosas sugestões.

REFERÊNCIAS

- BE PART OF THE WEB. Report of the Stanford Linked Data Workshop, 27 June – 1 July 2011. Disponível em: <<http://www.clir.org/pubs/abstract/pub152abst.html>>. Acesso em: 3 fev. 2012.
- BERMES, Emmanuelle. Convergence and Interoperability: a Linked Data perspective. In: IFLA World Library and Information Congress, 77th. Puerto Rico, 2011. Proceedings... 2011. Disponível em: <conference.ifla.org/past/ifla77/149-bermes-en.pdf>. Acesso em: 3 fev. 2012.
- BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web. Scientific American, May, 2001. Disponível em: <<http://www.scian.com/2001/0501issue/0501berners-lee.html>>. Acesso em: 24 maio 2001.
- BREITMAN, Karin. Web Semântica: a internet do futuro. Rio de Janeiro: LTC Editora, 2005.
- BYME, Gillian; GODDARD, Lisa. The Strongest Link: Libraries and Linked Data. Dlib Magazine, v.16 n. 11-12, Nov./Dec. 2010. Disponível em: <<http://www.dlib.org/dlib/november10/byrne/11byrne.html>>. Acesso em: 3 fev. 2012.
- BIZER, C.; CYGANIAK, R.; HEATH, T. How to publish Linked Data on the Web. [2007]. Disponível em: <<http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial/>>. Acesso em: 03 nov. 2011.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data – the story so far, In: T. Heath, M. Hepp, C. Bizer (eds.), Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS).

BOWEN, Jannifer B. Moving Library Metadata Toward Linked Data: Opportunities Provided by the eXtensible Catalog. In: International Conference on Dublin Core and Metadata Applications, Pittsburg, EUA, 2010, Proceedings... Disponível em: <<http://dcpapers.dublincore.org/ojs/pubs/article/view/1010>>. Acesso em: 28 abr. 2012.

DEFINITION OF THE EUROPEANA DATA MODEL ELEMENTS. Version 5.2.3, 24/02/2012. Disponível em: <<http://pro.europeana.eu/documents/900548/bb6b51df-ad11-4a78-8d8a-44cc41810f22>>. Acesso em 13 abr. 2012.

DUBLIN CORE METADATA ELEMENT SET. Version 1.1: Reference Description, 02 June 2003. Disponível em: <<http://dublincore.org>>. Acesso em: 29 jul. 2005.

EUROPEANA DATA MODEL PRIMER. ISAAC, Antoine (ed.). Europeana v 1.0. 2010. Disponível em: <<http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5>>. Acesso em: 3 fev. 2012.

THE EUROPEANA LINKED OPEN DATA PILOT. Disponível em: <<http://www.europeanaconnect.eu/news.php?area=News&pag=78>>. Acesso em: 3 fev. 2012.

FRBR – FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS : final report / IFLA Study Group on the Functional Requirements for Bibliographic Records. München: K. G. Saur, 1998. (UBCIM Publications New Series).

ISELE, Robert; Anja JENTZSCH, BIZER, Chris; VOLZ, Julius. Silk - A Link Discovery Framework for the Web of Data. October 6, 2010. Disponível em: <<http://www4.wiwiw.fu-berlin.de/bizer/silk/>>. Acesso em: abr. 20, 2012.

GUIZZARDI, Giancarlo. Ontological foundations for structural conceptual models. Enschede, The Netherlands: CTIT, Telematica Institut, 2005.

HANNEMANN, Jan. Linked Data for Libraries. In: WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY, 10-15 August 2010, Gothenburg, Sweden, Proceedings... Disponível em: <<http://www.ifla.org/files/hq/papers/ifla76/149-hannemann-en.pdf>>. Acesso em: 4 fev.2012.

LEGG, Catherine. Ontologies on the Semantic Web. *Annual Review of Information Science and Technology*, 2007, p. 407-451.

LIBRARY LINKED DATA INCUBATOR GROUP FINAL REPORT. W3C, 2011. Disponível em: <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>. Acesso em: 3 fev. 2012.

MILLER, Paul. Interoperability. What is it and why should I want it? *Ariadne*, v. 24, Jun. 2000. Disponível em: <<http://www.ariadne.ac.uk/issue24/interoperability/intro.html>>. Acesso em: 20 mar. 2002.

NILSONN, Mikael; POWEL, Andy; JOHNSTON, Pete; NAEVE, Ambjörn. Expressing Dublin Core metadata using the Resource Description Framework (RDF). 2010. Disponível em: <<http://dublincore.org/documents/dc-rdf/>>. Acesso em: 13 fev. 2012.

OWL Ontology Web Language Overview. 2004. Disponível em: <<http://www.w3.org/TR/owl-features/>>. Acesso em 15 mai. 2007.

RDA: Resource Description and Access. 2010. Disponível em: < <http://www.rda-jsc.org/rda.html>>. Acesso em 8 abr. 2012.

RDF PRIMER. MANOLA, Frank; MILLER, Eric (eds.). W3C, 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>>. Acesso em: 15 dez. 2011.

RFC 2396 - Uniform Resource Identifiers (URI): Generic Syntax, Berners-Lee T., Fielding R., Masinter L., IETF, August 1998. Disponível em: <<http://www.isi.edu/in-notes/rfc2396.txt>>. Acesso em: 15 dez. 2011.

RIVA, Pat; DOERR, Martin; ZUMER, Maja. FRBRoo: enabling a common view of information from memory institutions. In: WORLD LIBRARY AND INFORMATION CONGRESS: 74TH IFLA GENERAL CONFERENCE AND COUNCIL 10-14 August 2008, Québec, Canada, Proceedings. Disponível em: <<http://k.52caipu.com/newen/fl/iflanlc/iclc/IFLAds/201012/P020101210597174010207.pdf>>. Acesso em: 3 fev. 2012.

THE SEMANTIC WEB. Bulletin of The American Society for Information Science and Technology, v. 29, n. 4, Apr./May 2003. (Special Section).

SKOS. Simple Knowledge Organization System. 2004. Disponível em: <<http://www.w3.org/2004/02/skos/>>. Acesso em: 10 março 2012.

TAVARES, Luiz Fabiano de Freitas. Da Guanabara ao Sena: relatos e cartas sobre a França Antártica nas guerras de religião. Niterói: EDUFF, 2011.

W3C Library Linked Data Incubator Group. Datasets, Value Vocabularies, and Metadata Element Sets, 2011. Disponível em: <http://www.w3.org/2005/Incubator/lld/wiki/Vocabulary_and_Dataset>. Acesso em: 31 de março de 2012.