

Bibliotecas e instituições de memória na *web*, dados ligados e *web* semântica: diálogos interdisciplinares

Ana Lucia Pellegrino*
 Andrea Carvalho de Oliveira**
 Cláudio José Silva Ribeiro***
 Madalena Schmid Martins****
 Mariana Carvalho*****
 Suellen Alves da Silva*****
 Tiago Leite Pinto*****

Resumo

Na atualidade, a conexão entre as instituições de memória e seus usuários passa necessariamente pelo acesso à internet, o que representa a evolução dessas instituições no que tange à inovação na oferta de serviços a seus usuários. Este ensaio verifica a presença e a disponibilização de serviços de informação no ambiente *web* para bibliotecas, museus e arquivos, dialogando com a revisão de literatura sobre tecnologias e formalismos para representação da informação. Apoiar-se na interoperabilidade entre as informações disponíveis para viabilizar a melhoria dos serviços, bem como para agregar valor e dar maior visibilidade às instituições. Propõe representar conexões entre as informações disponibilizadas por essas instituições, usando formalismos e vocabulários específicos da *web* semântica. Esta narrativa é fruto da discussão desenvolvida na disciplina “Bibliotecas na *web*, dados ligados e *web* semântica”, no âmbito do mestrado profissional em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro, no primeiro semestre de 2016.

Palavras-chave: Biblioteca; Museu; Arquivo; Representação da informação; Web semântica; Dados ligados.

Abstract

Currently, the connection between memory institutions and their users necessarily passes through internet access, which represents the evolution of these institutions in what refers to

* Mestranda do Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (PPGB/Unirio) – apellegrino2005@gmail.com

** Mestranda do Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (PPGB/Unirio) – andreacoliveira@hotmail.com

*** Doutor em Ciência da Informação – Professor Adjunto do Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (PPGB/Unirio) – claudio.ribeiro@unirio.br

**** Analista de TIC – Bolsista Pesquisa de Desenvolvimento Científico e Tecnológico da Fundação Casa de Rui Barbosa (FCRB) – mschmidig@gmail.com.

***** Mestre em Biblioteconomia pelo Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (PPGB/Unirio) – carvmariana@gmail.com

***** Mestranda do Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (PPGB/Unirio) – suellenviegas@gmail.com

***** Mestrando do Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (PPGB/Unirio) – tiago.unirio@gmail.com

the innovation in the offer of services. This essay verifies the presence and availability of information services in the Web environment for libraries, museums, and archives, dialoguing with the literature review on technologies and formalisms for information representation. It relies on the interoperability between available information to allow the improvement of services, as well as adding value and giving greater visibility to these institutions. It proposes to represent connections between the information made available by them, using formalisms and specific vocabularies of the Semantic Web. This narrative is the result of the discussion developed in the discipline “Web Libraries, linked data, and Semantic Web”, given in the course of the professional master in Library Science of the Universidade Federal do Estado do Rio de Janeiro (Brazil), in the first semester of 2016.

Keywords Library; Museum; Archive; Information representation; Semantic Web; Linked data.

1 Introdução

Impulsionadas pelos temas “explosão informacional” e “avalanche de dados”, as instituições que lidam com memória e gestão de ativos de informação cultural têm buscado discutir a disponibilidade de seus acervos para além dos ambientes tradicionais de disseminação. Essa discussão passa pelo uso de vocabulários, formalismos e metainformação, pois estes componentes possuem um papel essencial no processo de representação e permitem desenvolver representações tanto no nível das ideias quanto no da sintaxe (RIBEIRO, 2014).

Com a exploração das propostas de representação já disponíveis para bibliotecas, museus e arquivos, surgem indagações que fomentam a discussão sobre a correlação de uma grande variedade de informações a partir da junção de conjuntos de dados publicados na Web. Enriquecer os dados com a inserção de novas dimensões para análise e disseminação é trabalho do profissional da informação, na medida em que é responsável por mediar o uso da informação nessas instituições.

Algumas discussões já foram encaminhadas nesse sentido, pavimentando o caminho para tratar a miríade de dados disponíveis na atualidade, especialmente no âmbito de instituições de memória. Marcondes (2012) promoveu o debate sobre dados ligados e sua contribuição para a *web* semântica. Seu trabalho pode ser incrementado com as visões colhidas em Ribeiro (2014) acerca do reflexo dos grandes volumes de dados nessas instituições, além da possibilidade de representar informação nesse novo cenário que se coloca para o cotidiano do profissional da área. Santos Neto e outros autores (2013) trazem um ensaio prático na representação de conexões, em que a possibilidade de descrever toda a sintaxe das ligações em código entendível por máquinas demonstra a viabilidade deste estudo.

É nesse sentido que este ensaio se apresenta, pois traz um breve diagnóstico sobre a presença na *web* das instituições de memória. Além disso, propõe representar ativos de informação cultural disponíveis em repositórios e busca interconectá-los com outros recursos de informação acessíveis *online*.

2 Panorama sobre instituições de memória na *web*

O estudo e o debate sobre as instituições de memória podem abordar vários aspectos e percepções. No âmbito dos objetivos deste ensaio, partimos para a análise da presença e da disponibilização de serviços de informação na *web* para bibliotecas, museus e arquivos, com o intuito de convalidar as visões do grupo que foram levantadas durante o desenvolvimento da nossa disciplina no Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro (PPGB/Unirio), no primeiro semestre de 2016.

Cabe destacar que a proposta de investigação nesses três tipos de instituições foi debatida e definida com o intuito de tornar as inspeções objetivas. Esse recorte tornou profícuo o processo de análise, principalmente por sabermos que, segundo a visão de memória da sociedade, as bibliotecas, os museus e os arquivos acabam por integrar seus acervos, simbolizando o conhecimento acumulado (DODEBEI, 2009).

2.1 Bibliotecas

As bibliotecas têm início com a necessidade de armazenamento da produção de conhecimento humano. Elas se relacionam com a descoberta da escrita em diversos tipos de materiais (SERRAI, 1975). São lugares de memória, assim como os arquivos e os museus, e, como eles, também dependem da produção e do registro documental para armazenar a memória, a fim de impedir que grande parte dela se perca com o passar do tempo, como ocorria antes da escrita (NORA, 2012). As bibliotecas e os suportes de escrita acompanharam os avanços da tecnologia e hoje temos bibliotecas digitais e virtuais, que são lugares de memória com amplo alcance.

Pode-se afirmar que a biblioteca digital abriga documentos digitalizados¹, organizados e armazenados segundo suas premissas. Compreendem a ideia de oferecer serviços e recursos para organizar e distribuir a informação, conservando a integridade desses documentos. Em

¹ O documento analógico, quando transformado em documento digital, tem seu conteúdo convertido em bits ou dígitos binários. Portanto, a noção de digitalização está ligada a essa transformação.

suma, uma biblioteca digital é uma coleção de serviços e objetos de informação, com organização, estrutura e apresentação específicas (BARRETO, 2010).

Quanto à biblioteca virtual, remete à noção de virtualização das bibliotecas tradicionais. Basicamente, se refere à ideia de uma biblioteca intangível, ou seja, um serviço de informação sem infraestrutura física. Assim, tem a vantagem de direcionar os usuários para as fontes de dados disponíveis no meio virtual, funcionando com o apoio da rede mundial, na qual são depositados diversos documentos, livros, monografias, imagens, vídeos, entre outros materiais (BARRETO, 2010).

Entretanto, com base na visão de Marchiori (2012), infere-se que os conceitos de biblioteca digital e virtual se confundem, já que as duas dependem do uso de tecnologias para existir:

Comumente, se reconhece como “bibliotecas digitais” (ou bibliotecas virtuais) as coleções de versões eletrônicas de um *corpus* de literatura publicada e tornada disponível através de uma rede interconectada de computadores, ou; como uma “poligamia” de documentos, tecnologia e serviços. (MARCHIORI, 2012).

Nesse sentido, destacamos não a diferença entre as duas, mas seus objetivos comuns. Ambas as propostas visam facilitar ao usuário o acesso à informação, disponibilizando elementos que possam ser de seu interesse. Marchiori continua sua reflexão e registra que as bibliotecas digitais armazenam objetos de informação como livros, periódicos, arquivos de som e imagem, entre outros, e oferece critérios de busca ligados ao título, autor e assunto. Esses tipos de biblioteca têm em comum o uso de ativos de informação em formato eletrônico ou sua descrição nesse formato. Toutain (2005) convalida nossa percepção quando nos diz se tratar de uma:

Biblioteca que tem como base informacional conteúdos em texto completo em formatos digitais – livros, periódicos, teses, imagens, vídeos e outros que estão armazenados e disponíveis para acesso, segundo processos padronizados, em servidores próprios ou distribuídos e acessados via rede de computadores em outras bibliotecas ou redes de bibliotecas da mesma natureza. (TOUTAIN, 2005).

Outro ponto de convergência é a facilidade na busca do usuário. Ambas têm em sua concepção serem utilizadas através da rede de computadores, interligadas, compartilhando informações para melhorar a disseminação destas.

Assim, pautados pela convergência entre as noções de biblioteca digital e virtual, bem como pelo recorte adotado para o ensaio, formulamos um breve diagnóstico sobre a presença na *web* de informações das bibliotecas públicas e das bibliotecas universitárias.

Para compreender melhor as primeiras, foram levantados dados sobre as bibliotecas municipais das 26 capitais dos estados brasileiros e do Distrito Federal. A análise considerou o sítio de cada biblioteca ou rede de bibliotecas, levando em conta informações sobre o *software* usado (quando era o caso), a forma de disponibilização do catálogo, a existência de acesso a algum acervo digital e observações consideradas importantes. Para facilitar a compreensão da amostra dos dados, a pesquisa foi dividida entre as cinco regiões do país, passando pelo Sudeste, Nordeste, Norte, Centro-Oeste e Sul. Por meio de pesquisa na *web*, foram identificadas as instituições com sítios disponíveis.

Em síntese, foram encontradas e visitadas quatro bibliotecas na região Sudeste, quatro no Nordeste e uma no Sul. Nas regiões Norte e Centro-Oeste nossa investigação não conseguiu encontrar e visitar nenhuma biblioteca pública municipal.

Com o intuito de ampliar o debate em curso na disciplina, decidiu-se por incrementar o escopo investigado, incorporando uma análise das bibliotecas universitárias. Essa escolha foi definida com base no *ranking Times Higher Education University*,² segundo o qual as cinco melhores universidades do Brasil são: Universidade de São Paulo (USP), Universidade Estadual de Campinas (Unicamp), Universidade Federal do Rio de Janeiro (UFRJ), Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) e Universidade Federal de Minas Gerais (UFMG).

Por meio de inspeção visual nos sítios, efetuada nos meses de maio e junho de 2016, verificou-se que as bibliotecas possuem muitas informações disponíveis para a comunidade usuária. Todas as universidades avaliadas disponibilizam seu acervo digital ou virtual, tanto o conteúdo quanto o catálogo de objetos. A possibilidade de acesso ao material impresso e digital e a um catálogo rico em detalhes sobre obras e localização foram elementos de destaque nos sítios das bibliotecas das cinco universidades.

2.2 *Museus*

Na visão de Nascimento (2010), um museu deve refletir de forma ampla a estrutura do tema a ser exibido. Além de cobrir aspectos de preservação, salvaguarda e gerenciamento de coleções, deve possibilitar também o melhor processo de comunicação e a respectiva compreensão de seu conteúdo (CERAVOLO, 2007).

²Disponível em: <<https://goo.gl/bOkDzK>>. Acesso em: 15 abr. 2017.

Em seu trabalho de investigação, Magaldi e Scheiner (2010) encaminham a discussão sobre a similaridade das noções de museu virtual, digital, *webmuseum* e *cibermuseu*. Adotaremos neste ensaio a visão de Carvalho (2012), segundo a qual o museu virtual pode ser entendido como um conjunto de objetos digitais que estão entrelaçados, para além dos métodos tradicionais de comunicar e interagir com os visitantes. Essa visão pode ser complementada por Oliveira (2007), quando este observa que o museu digital é o que possui presença *online*.

Os debates em sala de aula apontaram para o uso da publicação *Museus em Números*³ como fonte norteadora para o processo de busca e a realização da inspeção de dados. Apesar de essa ser uma extensa e completa edição sobre os museus do Brasil, não foi possível identificar informações sobre a informatização dos acervos e seu acesso *online*.

Procedeu-se, então, à busca de sítios, efetuada por meio de consultas utilizando o buscador Google com os parâmetros “museu”, “Brasil” seguidos da sigla da unidade da Federação.

Dos 43 sítios de instituições observadas, identificou-se que catorze possuem catálogo *online* com oferta de serviços para acesso aos registros. Foi na região Sudeste que obtivemos maiores resultados em buscas na *web*. Após a inspeção visual, apenas nove sítios foram categorizados como possuidores de oferta de serviços e/ou catálogos de bases de dados. Nas demais regiões, para a mesma categorização dos sítios, obtivemos os seguintes quantitativos: dois no Nordeste, um no Norte, um no Centro-Oeste e um na região Sul.

2.3 Arquivos

Os acervos arquivísticos brasileiros têm um leque de temas variados. Podemos encontrar arquivos pessoais, de música, literatura, de presidentes da República, manifestações culturais populares, documentos de povos tradicionais e de fatos históricos sobre países, como os acervos vinculados à Comissão Nacional da Verdade.

Como o objetivo da nossa investigação foi verificar a representação dos acervos arquivísticos em ambiente *web*, para analisar melhor o processo de inspeção foi necessário observar o sistema livre chamado ICA-AtoM. Este é um *software open source*⁴, criado em 2006 como uma iniciativa do Conselho Internacional de Arquivos (CIA) juntamente com

³ Disponível em <http://www.museus.gov.br/museus-em-numeros/>.

⁴ A modalidade *open source* (código aberto) caracteriza o conjunto de programas que estão disponíveis para a sociedade usar e alterar. São livres de custos de licenciamento.

outros parceiros, inclusive a Unesco. O ICA-AtoM é uma ferramenta que contempla as normas de descrição arquivística e possui um sistema multiarquivos e multirepositórios e também é um *software* multilíngue. Em sua versão mais recente (versão 2.0), passou a ser denominado AtoM (FLORES; HEDLUND, 2014).

O uso da solução ICA-AtoM foi observado no Arquivo Nacional (AN), no Diretório Brasil de Arquivos (Dibraq), na Base Arch da Fundação Oswaldo Cruz, no Arquivo Público de São Paulo, no Arquivo Público do Estado do Rio de Janeiro, no Acervo Arquivístico da Assembleia Legislativa do Estado de Minas Gerais, no Arquivo Digital da Universidade Federal do Rio Grande, no Arquivo Histórico Municipal João Spadari Adami, em Caxias do Sul, e na Universidade de Santa Maria.

O *DSpace*, outro *software* também disponível em modalidade *open source* para repositórios, é utilizado no Brasil por 101 instituições, dentre as quais temos estabelecimentos acadêmicos, governamentais, de pesquisa, de arquivo público, além de instituições cujo acervo é considerado múltiplo.⁵ Nessa lista, vale ressaltar a relevância, por exemplo, da Fundação Biblioteca Nacional, da Fundação Casa de Rui Barbosa (FCRB) e do Instituto Antonio Carlos Jobim.

Do total de cerca de dezessete instituições governamentais, a maioria contempla acervos históricos, classificados como obras raras, periódicos raros e arquivos históricos em geral.

Os arquivos históricos e institucionais da Fundação Casa de Rui Barbosa utilizam o *software* de catalogação para bibliotecas Sophia, como ocorre no Arquivo Museu de Literatura Brasileira (AMLB). Entretanto, parte das coleções digitalizadas empregam o *DSpace*, com link do *DocReader*⁶ para visualização; é o caso do Repositório Rui Barbosa de Informações Culturais (Rubi).

Na Fundação Casa de Rui Barbosa, a Diretoria de Pesquisa utiliza o *DSpace* integrado ao Portal de Preservação de Memória das Olimpíadas, com arquivos de depoimentos audiovisuais, artigos de jornais, entrevistas com atletas, mobilidade urbana, legados etc.

Vale ressaltar também que a Fundação Biblioteca Nacional recorre ao repositório *DSpace* para oferecer acesso e compartilhar suas coleções já digitalizadas, as quais são visualizadas por vários programas de imagem, como também pelo aplicativo *DocReader*.

Em relação aos arquivos públicos de memória e história do Brasil, foram pesquisadas 30 instituições e arquivos, perfazendo o total de 71 coleções. Em uma síntese sobre as

⁵ Disponível em: <http://duraspace.org/registry/>.

⁶ DOCPRO, disponível em: <https://goo.gl/5HfHbT>.

coleções visitadas, temos 61 na região Sudeste, três no Norte, três no Nordeste, uma na região Centro-Oeste e três no Sul do país.

3 Panorama sobre interoperabilidade e os formalismos para representação

Como visto anteriormente, os conceitos de biblioteca digital e virtual se confundem, pois as duas têm o objetivo de incrementar a disseminação do conhecimento por meio da maximização da busca e do uso da informação pelo usuário. Portanto, partindo-se deste pressuposto, pode-se afirmar que a utilização de estruturas tecnológicas por instituições de memória é essencial na propagação do conteúdo sob sua guarda. O amplo emprego de soluções em repositórios digitais potencializou nas instituições de memória a capacidade de armazenar, compartilhar e preservar conteúdo.

Nessa direção, um fator de grande relevância para essas instituições é a interoperabilidade, que terá a função importante de viabilizar a ligação entre elas, gerando aumento do acervo, além da facilidade de busca e recuperação da informação pelo usuário. A interoperabilidade depende não só de recursos tecnológicos, mas também de representação semântica, pois assim será possível resolver problemas complexos de significação, advindos do compartilhamento de informações entre repositórios de bibliotecas, museus e arquivos. É um processo que independe do *software* escolhido, mas depende de padrões estabelecidos para o espaço *web*. Para que as informações sejam utilizadas por diferentes instituições, com um amplo acervo compartilhado, é essencial que essas instituições estejam ligadas em rede e forneçam para o usuário um sistema de busca interoperável. Em suma, a interoperabilidade se traduz

[...] 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. (SAYÃO, 2007).

Sayão (2007b) diz que a interoperabilidade

[...] pode ser considerada como o processo contínuo de assegurar que sistemas, procedimentos e cultura de uma organização sejam gerenciados de modo a maximizar as oportunidades para intercâmbio e o reuso de informações. (SAYÃO, 2007b).

Assim, fica evidente a importância da interoperabilidade para as diferentes instituições de memória. Ela pode aumentar consideravelmente o acervo, bem como estender seu uso, facilitando o acesso pelo usuário. Para o funcionamento da interoperabilidade são determinantes fatores técnicos e semânticos, como protocolos, formalismos para representação e troca de informação (SAYÃO; MARCONDES, 2008).

Sayão e Marcondes (2008) continuam a reflexão observando exemplos de interoperabilidade técnica, como o protocolo Z39.50 e o formalismo XML, e de interoperabilidade semântica, que trata da representação da informação através de metadados, taxonomias, ontologias, entre outros recursos. No segundo tipo de interoperabilidade estão os formalismos RDF e OWL.

4 Web semântica, dados ligados e os formalismos para representação

A web semântica é formada por camadas de representação com linguagens e formalismos. Ela tem a intenção de atribuir sentido à informação contida na *web*, deslocando o uso sintático desta, mediado por humanos, na direção da contextualização e da possibilidade de mediação por não humanos. Promovendo maior interação entre usuários e máquinas, ela se utiliza de representações ontológicas para dar e correlacionar significados, gerando uma melhor busca na rede, ou seja, uma busca com resultados mais adequados, com melhora nos índices de revocação e precisão.

A correlação de dados e significados também é sustentada pelo conceito fundamental, formulado por Tim Berners-Lee, de *Linked Data* ou *Linked Open Data* (LOD), segundo o qual deve ser uma prática expor e compartilhar tanto os elementos de ligação de dados quanto as informações e o conhecimento. Essa plataforma tem como base o uso de URIs (*Universal/Uniform Resource Identifier*) para identificação e endereçamento de unidades de informação, além de RDF (*Resource Description Framework*) para representá-las.⁷

Para Marcondes (2012), a estruturação do espaço *web* não se fará apenas por meio de *links* entre documentos, mas sim com uma nova abordagem em que documentos e dados

⁷ BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. *Linked Data: the story so far*.

passem a ser ligados, recorrendo a um conjunto de melhores práticas para publicá-los e torná-los disponíveis. Essa proposta possui grande potencial, pois conecta ativos de informação por meio de *links* semânticos que também permitem interpretação por programas computacionais. Essa interpretação se dará por meio da representação gerada com o uso de linguagens e formalismos específicos.

Na visão de Ribeiro (2015), os formalismos para representação podem ser classificados segundo seus níveis de abstração: formalismos que auxiliam a descrever a representação no nível verbal da sintaxe (nível lógico textual) e formalismos que operam no nível das ideias (nível lógico dos nossos pensamentos e inferências). O recorte adotado por este ensaio procurou elucidar apenas os formalismos utilizados para descrever os elementos alvo de representação – SKOS, FRBR, DC, OWL e RDF – não esgotando todas as alternativas disponíveis, tanto para o nível da abstração verbal quanto para o das ideias.

4.1 SKOS

Simple Knowledge Organization System é um padrão de sistemas de organização do conhecimento recomendado pelo W3C (*World Wide Web Consortium*), que considera a organização de conteúdos informacionais na *web* (RAMALHO, 2015). Ramalho, Vidotti e Fujita (2007) corroboram a definição, afirmando que

O W3C tem se empenhado no desenvolvimento de padrões que deem suporte ao uso de Knowledge Organization Systems, KOS a partir da estrutura da Web Semântica, entre os quais pode-se destacar o *Simple Knowledge Organisation System* (RAMALHO; VIDOTTI; FUJITA, 2007).

O SKOS representa desde estruturas simples de sistemas de organização da informação até os mais complexos. Fornecendo estrutura básica para representar esquemas conceituais e ontologias, também pode ser combinado com outros formalismos, como OWL (RAMALHO; VIDOTTI; FUJITA, 2007). Ribeiro (2015) complementa caracterizando o SKOS como um vocabulário que possibilita a representação semiformal de tesouros, taxonomias e esquemas de classificação no ambiente *web*.

A representação se dá por meio de dois tipos de propriedade, que descrevem formalmente os conceitos: as propriedades de etiquetagem e as de documentação. As primeiras permitem o uso de rótulos (*labels*) para a descrição de aspectos específicos, ou seja, elas possibilitam o controle da terminologia utilizada para os conceitos. Já as propriedades de documentação, têm a função de documentar os conceitos, incluindo suas características

informais e reunindo informações importantes dos esquemas representados (RAMALHO, 2007).

O SKOS tem uma forte ligação com formalismos como RDF e OWL, o que facilita a troca de informações, a interoperabilidade entre sistemas e a reutilização de dados. Seu uso representa uma boa relação entre signos utilizados na representação do conhecimento (RIBEIRO, 2015).

4.2 FRBR

O *Functional Requirements for Bibliographic Records* é um modelo conceitual para representação de registros bibliográficos baseado em requisitos mínimos e na relação entidade-relacionamento (RIBEIRO, 2015). Maimone, Silveira e Tálamo (2011) explicam que “nos modelos conceituais do tipo entidade-relacionamento são apresentadas as entidades, seus atributos e seus relacionamentos”. Melo e Bräscher (2014) incluem nessa definição a ideia de que o modelo FRBR foi criado com

[...] o objetivo de ampliar o escopo da representação da informação, ao contemplar as relações entre determinados elementos de dados do registro bibliográfico e considerar os questionamentos dos usuários que consultam essas informações, numa abordagem conceitual do universo bibliográfico e independente de um código de catalogação específico.

O vocabulário que expressa o modelo FRBR em linguagem processável por máquinas permite que a recuperação da informação seja mais efetiva. Isto decorre da compatibilidade com outros sistemas de informação e com tipos de representação como o RDF e o XML. Esse formalismo para descrição na *web* é composto por 69 elementos que estão apoiados por classes RDF, contemplando os grupos 1, 2 e 3 do modelo FRBR (RIBEIRO, 2015).

4.3 DC

O *Dublin Core* foi originalmente concebido para gerar descrições de recursos na *web* por isso atraiu a atenção de bibliotecas e museus (WEIBEL, 2005). Para Alves e Souza (2007), o DC é um padrão de metadados que permite a descrição de recursos eletrônicos, com possibilidade de estender seus elementos para cobrir novas necessidades de descrição.

O formalismo DC também usa o RDF em sua sintaxe e é composto por quinze elementos considerados como núcleo da proposta. Por utilizar elementos com estrutura XML,

permite a extensão e criação de novos elementos de maneira muito simples, tornando a proposta de representação bastante flexível (RIBEIRO, 2015).

4.4 OWL

A *Web Ontology Language* é um formalismo que possibilita definir ontologias na *web*. Segundo a W3C (2004), OWL “é projetado para uso por aplicativos que precisam processar o conteúdo da informação em vez de apenas apresentar informações para os seres humanos.” É suportado pelo RDF na descrição de ontologias complexas na web semântica, com alto nível de expressividade em suas sentenças.

Ainda segundo a W3C (2004), OWL acrescenta mais vocabulário para descrever as propriedades e classes, ou seja, descreve as relações entre propriedades e classes. Estas relações também podem ser formalmente modeladas com base na lógica descritiva. OWL expressa a intenção de modelar do investigador, pois é possível registrar construções complexas, tais como propriedades encadeadas ou ainda a restrição entre as classes e valores, ou ainda em propriedades e hierarquias (RIBEIRO, 2015).

4.5 RDF

Para W3C (2014), a *Resource Description Framework* fornece uma descrição semântica para a relação dos objetos e seus recursos. (, ou seja, para uma descrição das características a descrição dos recursos sobre os objetos). Ela tem uma boa representação para o tratamento de metadados e é composta pela trinca: sujeito – predicado – objeto. É um formalismo que descreve recursos de informação, apoiado por uma arquitetura de metadados voltada para a descrição de recursos na *web*. Expressa em linguagem XML, a RDF possui a característica de permitir o processamento e a interpretação por máquinas. Seu traço marcante é a possibilidade de assumir o papel de gramática comum às linguagens de representação (RIBEIRO, 2015).

Como registrado nos formalismos apresentados anteriormente, os elementos descritos neste ensaio possuem uma forte correlação com a RDF, o que permite que todos os formalismos adotem uma gramática descritiva comum. Adicionalmente, é possível também executar a representação por meio de recursos gráficos (grafos RDF), o que pode instrumentar melhor as discussões sobre o tema (RDF, 2014).

5 Metodologia

Neste ensaio, partiu-se do pressuposto de que as bibliotecas, os museus e os arquivos tratam da memória coletiva como reservatórios ricos em documentos (LE GOFF, 1990). Além disso, essas instituições também contribuem para a formação da memória social, abordando os problemas do tempo e da história (DODEBEI, 2009). Portanto, devem possibilitar que o uso das novas tecnologias contribua para o seu desenvolvimento. Como instituições detentoras e disseminadoras da informação, é essencial que estas tenham seus acervos compartilhados, permitindo a criação de laços entre informações e de novos conhecimentos.

A revisão de literatura abarcou definições sobre as instituições de memória, tecnologias e formalismos para representação da informação. A partir disso, foram pesquisados os endereços eletrônicos das instituições e inspecionados seus respectivos sítios na *web*, em busca de informações sobre disponibilização de acervo, existência de catálogos, uso de padrões para representação do acervo *online*, além da disponibilidade do conteúdo digital por completo.

Foram gerados quadros intermediários e comparativos entre os sítios inspecionados, sistematizando-se a análise por meio da agregação das informações em categorias. Para isso, considerou-se especialmente a disponibilização do acervo digital, tanto em conteúdo quanto em referências.

Com o diagnóstico formulado, fruto das discussões desenvolvidas no bojo da já citada disciplina, foram definidos requisitos para o ensaio de representação, a saber:

- Uso de ativos de informação com conteúdo em acervo digital;
- Uso de ativos de informação ligados a personalidades com projeção mundial;
- Uso de ativos de informação que permitissem conexões com outros contextos.

Nessa direção, identificamos o arquivo de Rui Barbosa como adequado para o nosso ensaio. A escolha da coleção foi determinada por debates em sala de aula, levando em conta o conhecimento e a experiência de cada profissional envolvido. O recorte da pesquisa contemplou a Campanha Civilista, pois esta pode ser considerada como a primeira campanha presidencial que contou com a participação de diversos segmentos da sociedade brasileira. Nela, o candidato Rui Barbosa concorreu com o candidato militar apoiado pelo governo – Hermes da Fonseca. Além de representar a luta pela consolidação da ordem civil no Brasil, a

Campanha demonstrou o desejo da sociedade por uma participação política mais efetiva, em contraposição ao domínio político de tradicionais oligarquias.

A representação da coleção foi incrementada a partir de sugestões de laços entre informações passíveis de correlação. Ao final, foi gerada uma representação gráfica em formato RDF Graph, com a representação de trincas RDF (sujeito, predicado e objeto).

6 Resultados e discussão

Foi possível perceber, na investigação, que a maioria das instituições disponibiliza algum tipo de instrumento para identificar seus acervos, podendo ser uma base de dados ou um catálogo de referências. No entanto, nem sempre essas instituições proporcionam acesso a objetos digitais. Nesse sentido, constatou-se a diferença entre sítios informativos sobre a instituição e sítios que oferecem algum tipo de serviço para acesso ao conteúdo.

6.1 Bibliotecas

Verificou-se que, das 26 bibliotecas municipais das capitais brasileira analisadas e do Distrito Federal, apenas nove oferecem serviços com informações sobre seu catálogo e acesso ao acervo.

Já para bibliotecas universitárias, o diagnóstico foi mais favorável. Por serem instituições de ensino e pesquisa, é esperado que seus acervos estejam disponíveis para a sociedade em geral. Entretanto, apesar de os acervos estarem disponíveis, notou-se que a temática tratada neste ensaio ainda carece de investimentos por parte das universidades observadas. Em síntese, as cinco instituições analisadas possuem acervo digital, mas sem oferecer alternativas para conexões de seu conteúdo com outros sítios na *web*.

Assim, a análise dos sítios que tratam o conhecimento bibliográfico acumulado permitiu deduzir que, no âmbito das bibliotecas universitárias o esforço para implementação da interconexão será menor que no das municipais, em decorrência da pequena presença destas na *web*.

6.2 Museus e Arquivos

Após debates em sala de aula, decidiu-se apresentar os resultados ligados a museus e arquivos em conjunto, pois na nossa investigação foram identificados espaços na *web* que abrigam coleções arquivísticas, mas que acabam também sendo enquadrados nas definições de museu, conforme apresentado anteriormente, na seção 2.2. Vale destacar, ainda, que o ferramental tecnológico observado nas inspeções também apresentou características que contribuíram para uma análise conjunta das instituições.

Para museus, dos 43 sítios inspecionados, a existência de catálogo *online* com alguma oferta de serviços esteve presente em catorze.

Nos trinta sítios de arquivos inspecionados foram visitadas 71 coleções. O processo de análise passou a tratar essas coleções como base de comparação, e não mais a instituição, pois o mesmo sítio pode prover acesso ao conteúdo de uma coleção e para a outra simplesmente apresentar dados informativos. Das 71 coleções inspecionadas, identificou-se que 48 têm catálogo *online* com serviços para acesso aos registros.

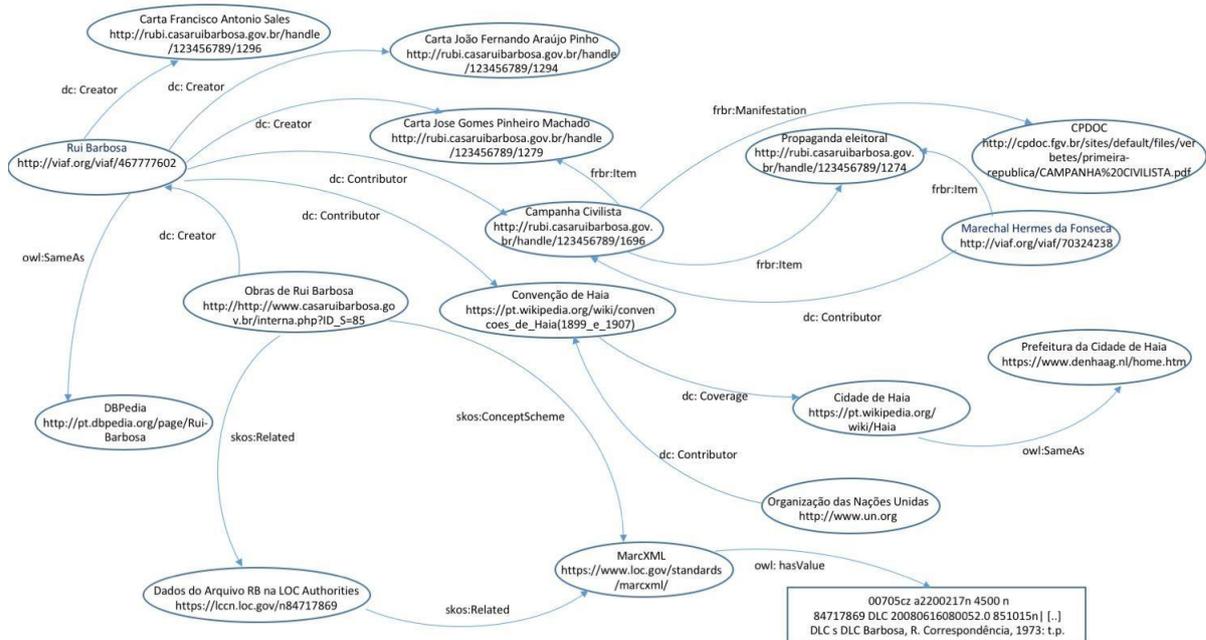
Assim, partindo do pressuposto de que instituições e/ou coleções objetivam armazenar o conhecimento não bibliográfico acumulado, pode-se fazer uma quantificação que nos auxiliou no diagnóstico. Verificou-se que destes 114 componentes de registro de memória da sociedade inspecionados, 43 de museu e 71 de arquivo, cerca de 54% está disponibilizando serviços para acesso ao seu conteúdo.

Adicionalmente, o processo de inspeção permitiu diagnosticar também outro aspecto considerado relevante pelos autores: o uso de *software* proprietário para acesso a coleções. Em contraposição ao *software open source*, tratado anteriormente, esse tipo de *software* é desenvolvido por uma ou mais empresas que adotam como estratégia comercial os custos de licenciamento. Tal solução foi identificada em cerca de 39% das coleções inspecionadas. Além de demandar recursos financeiros para sua aquisição, o *software* proprietário pode acarretar alguma dificuldade no alinhamento com padrões para interconexão de objetos digitais com outros ambientes, pois o alinhamento a esses padrões precisa ser implementado pelo proprietário do aplicativo. Segundo as regras que norteiam o mercado de *software*, deve-se ter algum investimento adicional associado à implementação dessa interoperabilidade e isso pode inviabilizar ações para interconexão.

A partir do diagnóstico presente neste trabalho, verificou-se que no Brasil a organização e estruturação de arquivos e museus para consultas na *web* ainda carece de evolução, especialmente no âmbito de serviços para interconexão com outros ambientes.

Assim, com o intuito de incrementar as pesquisas em representação para as instituições de memória, em especial para grandes coleções que começam a estar presentes em ambientes digitais, foi formulada a proposta de representação, fruto das discussões, conforme vemos na Figura 1:

Figura 1 – Grafo RDF⁸ sobre Rui Barbosa e cartas da Campanha Civilista



Fonte: os autores.

Por outro lado, os resultados obtidos na análise dos serviços das bibliotecas investigadas, especialmente no âmbito das universidades, demonstram que estas estão um pouco à frente, principalmente no tocante ao compartilhamento de informações. O uso e a disponibilização de catálogos de referência, aliados à preocupação frequente com a coleta e troca de metadados entre instituições, nos leva a inferir que o caminho a ser trilhado por essas bibliotecas está um pouco melhor pavimentado, ou seja, tudo leva a crer que caminham para fazê-lo num futuro próximo. No entanto, a representação gerada por este ensaio poderá servir de estímulo para o incremento de estudos na direção de disseminar seus acervos.

⁸Como não é objetivo deste relato apresentar o *framework* RDF, seguem algumas recomendações para realizar a leitura do Grafo: o conceito embutido na representação é sempre a trinca sujeito – predicado – objeto. A leitura deve ser feita sempre no sentido da seta; o predicado é sempre o rótulo que está junto à seta. Por exemplo, pode-se ler na representação que Rui Barbosa (sujeito) criou (predicado) a carta para José Gomes Pinheiro Machado (objeto). Cabe registrar adicionalmente que, dependendo da ótica da representação, o sujeito e o objeto podem trocar de posição.

7 Considerações finais

A visão da *web* semântica é essencial para conseguirmos tratar o alto volume de dados e informação que está disponível na grande rede. Publicar e consumir informações na atualidade é fator corriqueiro para todos nós, quer sejamos usuários da sociedade em geral, com nossos celulares e *tablets*, quer sejamos pesquisadores especializados, em estações de trabalho procurando novas informações e correlações. Nesse sentido, ao fazer uma busca é preciso obter resultados diferentes de números soltos e palavras sem contexto. O uso de mecanismos de busca automatizados permite que as pesquisas possam ser desenvolvidas de forma mais rápida, auxiliando na análise e filtragem de resultados.

Por exemplo, com a transcrição do grafo RDF da Figura 1 em linguagem processável por máquinas (RDF/XML) e implementação das ligações no modelo proposto, será possível, ao realizar uma pesquisa na *web*, que um robô automatizado (mecanismo ou, em inglês, *engine*) identifique que Rui Barbosa e Hermes da Fonseca estiveram presentes na Campanha Civilista, inclusive com a disponibilização de imagens representativas da propaganda eleitoral.

Indo nessa direção, a representação e disponibilização de outras coleções seguindo esses formalismos permitirão um alto grau de interconexão entre dados e informação, tornando a *web* cada vez mais próxima do conceito de teia, mas agora com ligações dotadas de valor semântico.

Para isso, a utilização de formalismos e abordagens para criar ligações entre dados e informação é mister, pois pode impulsionar as investigações em busca da noção de humanidades digitais. Na visão de Dacos (2011), na contemporaneidade ocorrem alterações na produção e na disseminação de conhecimentos a partir da opção da sociedade por trilhar os caminhos do mundo digital.

Referências

ALVES, Maria das Dores Rosa; SOUZA, Marcia Isabel Fugisawa. Estudo de correspondência de elementos metadados: Dublin Core e Marc 21. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 4, n. 2, p. 20-38, jan./jun. 2007.

BARRETO, Aldo. **Bases de dados, repositórios de informação, bibliotecas digitais e virtuais**. 21 abr. 2010. Disponível em: <<https://goo.gl/X09JPH>>. Acesso em: 1º abr. 2016.

BIZER, Christian; HEATH, Tom; BERNERS-LEE, Tim. **Linked data: the story so far**. Disponível em: <<https://goo.gl/fvnecy>>. Acesso em: 29 jun. 2016.

CARVALHO, Rosane Maria Rocha de. Comunicação e informação de museus na internet e o visitante virtual. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. **Anais...** Rio de Janeiro: Questões em Rede, 2012. Disponível em: <<https://goo.gl/dEAuya>>. Acesso em: 1º mar. 2017.

CERAVOLO, Suely Moraes; TÁLAMO, Maria de Fátima. Os museus e a representação do conhecimento: uma retrospectiva sobre a documentação em museus e o processamento da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais...** Salvador: Enancib, 2007.

DACOS, Marin. Manifeste des Digital Humanities. In: THATCAMP PARIS, 2011, Paris. **Anais...** Paris: THATCamp, 2011. Disponível em: <<https://goo.gl/m5UBtQ>>. Acesso em: jan. 2017.

DOCPRO. **Ambiente**. [2016?]. Disponível em: <<https://goo.gl/5HfHbT>>. Acesso em: 15 abr. 2017.

DODEBEI, Vera. Repositórios institucionais: por uma memória criativa no ciberespaço. In: SAYÃO, Luís Fernando et. al. (Orgs.). **Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: Edufba, 2009. p. 83-106.

DURASPACE registry. 2016. Disponível em: <<https://goo.gl/0yIBF2>> Acesso em: 1º mar. 2017.

FLORES, Daniel; HEDLUND, Dhion Carlos. Análise e aplicação do Ica-Atom como ferramenta para descrição e acesso às informações do patrimônio documental e histórico do município de Santa Maria – RS. **Informação & Informação**, Londrina, v. 19, n. 3, p. 86-106, set./dez. 2014.

LE GOFF, Jacques. **História e memória**. Tradução de Bernardo Leitão et al. Campinas: Unicamp, 1990.

MAGALDI, Monique Batista; SCHEINER, Tereza Cristina. Reflexões sobre o museu virtual. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: INOVAÇÃO E INCLUSÃO SOCIAL: QUESTÕES CONTEMPORÂNEAS DA INFORMAÇÃO, 11., 2010, Rio de Janeiro. **Anais...** Rio de Janeiro: Enancib, 2010. Disponível em: <<https://goo.gl/NBDhzb>>. Acesso em: 15 abr. 2017.

MAIMONE, Giovana Deliberali; SILVEIRA, Naira Christofolletti; TÁLAMO, Maria de Fátima. Reflexões acerca das relações entre representação temática e descritiva. **Informação & Sociedade: Estudos**, João Pessoa, v. 21, n. 1, p. 27-35, jan./abr. 2011. Disponível em: <<https://goo.gl/9Jl8Lm>>. Acesso em: 29 jun. 2016.

MARCHIORI, Patrícia Zeni. Bibliotecas digitais e repositórios de objetos de aprendizagem. **Informação & Sociedade: Estudos**, João Pessoa, v. 22, n. 2, p. 13-21, 2012.

MARCONDES, Carlos Henrique. Linked data: dados interligados e interoperabilidade entre arquivos, bibliotecas e museus na web. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 17, n. 34, p. 171-192, maio/ago. 2012. DOI: 10.5007/1518-2924.2012v17n34p171.

MELO, Maria Antônia Fonseca; BRÄSCHER, Marisa. Requisitos funcionais para dados de autoridade assunto (FRSAD): entidades, atributos e relacionamentos. **Revista digital de biblioteconomia & ciência da informação**, Campinas, v. 12, n. 2, p. 102-119, 2014. Disponível em: <<https://goo.gl/pq3MXb>>. Acesso em: 29 jun. 2016.

MUSEUS em números. Brasília, DF: Instituto Brasileiro de Museus, 2011. v. 1. Disponível em: <<https://goo.gl/oat9z3>>. Acesso em: 15 abr. 2017.

NASCIMENTO, Tereza Cristina Moraes do. Museus Virtuais de Ciências: historicizando o conceito de museu, popularização da ciência e inclusão digital. **Revista Museu Virtual**, Rio de Janeiro, v. 1, n. 1, p. 12-24, jan./dez. 2010. Disponível em: <<https://goo.gl/ehBW3n>>. Acesso em: 1º abr. 2017.

NORA, Pierre; AUN KHOURY, Tradução: Yara. Entre Memória e História: a problemática dos lugares. Projeto História. **Revista do Programa de Estudos Pós-Graduados de História**, [S.l.], v. 10, out. 2012. ISSN 2176-2767. Disponível em: <<https://revistas.pucsp.br/index.php/revph/article/view/12101/8763>>. Acesso em: 15 abr. 2017.

OLIVEIRA, José Cláudio. O museu digital: uma metáfora do concreto ao digital. **Comunicação e Sociedade**, Braga, v. 12, p. 147-161, 2007. Disponível em: <<https://goo.gl/4rZbJ1>>. Acesso em: 15 abr. 2017.

OWL WEB ONTOLOGY LANGUAGE OVERVIEW. 2004. Disponível em: <<https://goo.gl/H8c7On>>. Acesso em: 28 mar. 2016.

RAMALHO, Rogério Aparecido Sá. Análise do modelo de dados Skos: sistema de organização do conhecimento simples para a web. **Informação & Tecnologia**, Marília/João Pessoa, v. 2, n. 1, p. 66-79, 2015. Disponível em: <<http://bit.ly/2roybbg>>. Acesso em: 29 jun. 2016.

_____; VIDOTTI, Silvana Aparecida Borsetti Gregório; FUJITA, Mariângela Spotti Lopes. Web semântica: uma investigação sob o olhar da ciência da informação. **DataGramZero**: revista de ciência da informação, João Pessoa, v. 8, n. 6, dez. 2007. Disponível em: <<https://goo.gl/JQyKbd>>. Acesso em: 26 mar. 2016.

RDF – RESOURCE DESCRIPTION FRAMEWORK. 25 fev. 2014. Disponível em: <<https://goo.gl/HI8tMF>>. Acesso em: 27 mar. 2016.

RIBEIRO, Cláudio José Silva. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia (ITEC)**, Marília/João Pessoa, v. 1, n. 1, p. 96-105, 2014.

_____. Organização da informação no ambiente Web: produzindo conhecimento a partir de grandes volumes de dados. In: Congreso ISKO España, 12.; Congreso ISKO España-Portugal, 2., 2015, Murcia. **Actas...** Murcia: Isko, 2015.

_____. Uma investigação sobre o uso de vocabulários e formalismos: em busca de um caminho para a representação semântica de patrimônio digital e ativos de informação cultural. In: DESAFÍOS Y OPORTUNIDADES DE LAS CIENCIAS DE LA INFORMACIÓN Y LA DOCUMENTACIÓN EN LA ERA DIGITAL: Encuentro Iberico EDICIC, 7. 2015, Madrid. **Actas...** Madrid: Universidad Complutense de Madrid, 2015.

SANTOS NETO, Antônio Laurindo dos et. al. Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. **TransInformação**, Campinas, v. 25, n. 1, p. 81-87, jan./abr., 2013.

SAYÃO, Luís Fernando. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes – URN, PURL, DOI, Handle System, CrossRef e OpenURL. **Transinformação**, Campinas, v. 19, n. 1, p. 65-82, jan./abr. 2007. Disponível em: <<https://goo.gl/BMYj1q>>. Acesso em: 16 jun. 2015.

_____. Padrões para bibliotecas digitais abertas e interoperáveis. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 12, p. 18-47, 1. sem. 2007. Número especial. Disponível em: <<https://goo.gl/5b7MmI>>. Acesso em: 1º abr. 2016.

SAYÃO, Luís Fernando; MARCONDES, Carlos Henrique. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**, Campinas, v. 20, n. 2, p. 133-148, maio/ago. 2008. Disponível em: <<https://goo.gl/z7AWxm>>. Acesso em: 4 maio 2015.

SERRAI, A. História da biblioteca como evolução de uma idéia e de um sistema. **R. Esc. Bibliotecon.** UFMG, Belo Horizonte, v. 4, n. 2, p. 141-161, set. 1975.

SILVA, Renata Eleuterio da; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Requisitos funcionais para registros bibliográficos (FRBR): considerações sobre o modelo e sua implementabilidade. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 8, n. 2, p. 116-129, 2012. Disponível em: <<https://goo.gl/7MzrQS>>. Acesso em: 29 jun. 2016.

TOUTAIN, Lúcia Maria Batista Brandão. Biblioteca digital: definição de termos. In: SAYÃO, Luís Fernando et al. (Orgs.) **Bibliotecas digitais: saberes e práticas**. Salvador: Edufba, 2005. p. 15-24.

WEIBEL, Stuart. The Dublin Core: a sample content description model for electronic resources. **Bulletin of the American Society for Information Science**, Silver Spring, v. 24, n. 1, p. 9-11, 2005.