

ARQUITETURA PARA A DESCRIÇÃO BIBLIOGRÁFICA SEMÂNTICA: requisito para a interoperabilidade em ambientes informacionais digitais

Fabiano Ferreira de Castro

Doutor em Ciência da Informação, Coordenador do Curso de Biblioteconomia da Universidade Federal de Sergipe – UFS/SE

Plácida Leopoldina Ventura Amorim da Costa Santos

Livre-docente em Catalogação, Professora da Universidade Estadual Paulista – UNESP/Marília

RESUMO

Nos ambientes informacionais digitais atualmente, onde vários tipos de recursos coexistem com formatos e padrões de metadados heterogêneos, muitos esforços têm sido feitos para alcançar a interoperabilidade, a fim de utilizar vários padrões de metadados e de reutilizar os registros de metadados, através do desenvolvimento de estratégias, que vão desde os simples mapeamentos entre elementos de metadados à modelagem estrutural complexa. O tratamento dado ao recurso informacional requer uma descrição de forma e de conteúdo legível por máquinas, com resultados compreensíveis aos humanos, e, que ainda, atenda aos requisitos de interoperabilidade entre os ambientes informacionais. Nesse sentido, a investigação propõe a reflexão e um olhar pautados numa proposta de desenvolvimento de uma arquitetura para a descrição bibliográfica semântica, que garanta a interoperabilidade em ambientes informacionais digitais. Por meio de uma metodologia de análise exploratória e descritiva do tema, verifica-se que as metodologias da Catalogação Descritiva e uma ontologia para a descrição bibliográfica, explicitada nas regras e códigos de catalogação, e nos padrões de metadados redesenham a construção de novos ambientes informacionais digitais melhor estruturados, para a recuperação da informação e no estabelecimento efetivo da interoperabilidade.

Palavras-chave: Interoperabilidade. Padrões de metadados. Modelagem de ambientes informacionais digitais. Catalogação automatizada. Ontologia para descrição bibliográfica. Informação e Tecnologia.

ARCHITECTURE FOR THE SEMANTIC BIBLIOGRAPHIC DESCRIPTION: requirement for interoperability in digital informational environments

ABSTRACT

Actually in the digital informational environments, where several types of resources coexist with shapes and patterns of heterogeneous metadata, many efforts have been made to achieve interoperability in order to use multiple metadata patterns and reuse the metadata records through the development of strategies, ranging from simple mappings between metadata elements to the complex structural modeling. The treatment of the information resource

requires a description of form and content machine readable, understandable results to humans, and which also meets the requirements of interoperability between information environments. In this sense, the research proposes a reflection and a look lined a proposal to develop an architecture for the bibliographic description semantics, which ensures interoperability in digital informational environments. Through an analysis methodology exploratory and descriptive of the subject, it appears that the methodologies of Descriptive Cataloging and ontology for bibliographic description, explained the rules and codes of cataloging and metadata standards to redesign the construction of new digital informational environments better structured for information retrieval and effective establishment of interoperability.

Keywords: Interoperability. Metadata patterns. Modeling of digital informational environments. Automatized cataloguing. Ontology for bibliographic description. Information and Technology.

1 INTRODUÇÃO

Pode-se observar contemporaneamente um cenário marcado por grandes transformações tecnológicas e pelo surgimento e criação das mais variadas técnicas para as questões de representação, de descrição, de produção, de organização, de transmissão e de uso das informações em ambientes informacionais digitais.

O ambiente *Web*, por possuir informações heterogêneas e efêmeras, pode ser caracterizado como caótico, onde profissionais de diversas áreas buscam soluções para um tratamento mais significativo dos conteúdos que ali estão armazenados e disponibilizados para um uso mais efetivo. Idealizada por Tim Berners-Lee e liderada pelo *World Wide Web Consortium* (W3C) - a *Web Semântica* é um projeto que visa a solucionar e a minimizar os problemas de recuperação da informação na *Web* por meio do acesso automatizado aos recursos informacionais. Para tanto, pretende implantar uma estrutura tecnológica e instaurar uma maior representação do conhecimento na rede com o estabelecimento semântico das informações contidas nas unidades e assim, possibilitar técnicas mais eficazes de recuperação (BERNERS-LEE; HENDLER; LASSILA, 2001, tradução nossa).

Durante as últimas décadas observou-se que mais e mais informações têm sido publicadas, armazenadas e disponibilizadas na *Web*. No entanto, somente uma parte dessas informações e em ambientes específicos, tais como as bibliotecas digitais, por exemplo, possuem um padrão de metadados para a descrição bibliográfica e que possibilite a padronização dos recursos informacionais para atender aos requisitos da interoperabilidade.

As bibliotecas digitais apresentam-se e atuam como um segmento na Internet onde procuram desenvolver e criar métodos e técnicas para a padronização dos recursos informacionais. Mas garantir/assegurar a interoperabilidade entre “ilhas” de informações bem estruturadas e padronizadas entre formatos bibliográficos distintos encontra-se numa questão de investigação e preocupação pela comunidade científica. (CASTRO; SANTOS, 2010).

Novas tecnologias baseadas em pesquisas na temática *Web Semântica* desenvolvem-se com o intuito de tentar agregar semântica à descrição bibliográfica para um tratamento mais efetivo dos conteúdos digitais. A *Web Semântica* oferece algumas soluções para diferentes problemas criando uma nova visão para o armazenamento e o processamento dos dados. Algumas dessas soluções poderiam ser implementadas para potencializar os resultados das buscas no âmbito das bibliotecas digitais.

A preocupação da comunidade biblioteconômica advinda com as novas regras para a descrição bibliográfica aplicadas no ambiente digital, levam os profissionais a repensarem o

seu fazer profissional, na tentativa de caminharem em consonância com as novas transformações trazidas no bojo da recontextualização tecnológica das bibliotecas.

Dessa forma, pretende-se apontar e refletir sobre tais mudanças, numa proposta de um modelo de requisitos funcionais, para a modelagem de catálogos bibliográficos, pautados na lógica descritiva das metodologias da Catalogação Descritiva, nas ontologias para a descrição bibliográfica explicitadas nas regras e códigos de catalogação e em padrões de metadados, no estabelecimento efetivo da interoperabilidade em ambientes informacionais digitais.

2 A CATALOGAÇÃO DESCRITIVA NO SÉCULO XXI: impactos no ambiente digital

No cenário atual da comunidade da Biblioteconomia e da Ciência da Informação, há uma preocupação e um reconhecimento crescente de que haverá a necessidade de um sucessor do formato bibliográfico MARC 21, devido às novas transformações no domínio bibliográfico, permeada pelo uso intensivo das Tecnologias de Informação e Comunicação (TIC). Segundo Coyle (2011) tais discussões tendem a se concentrar principalmente nas questões estruturais, o novo formato será XML (*eXtensible Markup Language*), ele fará uso de RDF (*Resource Description Framework*) e padrões de dados ligados¹?

O que essas questões não se referem é a tarefa muito mais complexa de traduzir a semântica dos dados bibliográficos para a construção de um novo padrão. De acordo com Thomale (2010) basta apenas uma pequena investigação dos dados codificados no formato MARC 21, para revelar que as *tags* e os subcampos em si, são insuficientes para definir os elementos dos dados reais realizados pelos registros de um catálogo bibliográfico. “O primeiro passo na transformação de MARC 21 para um outro formato é identificar quais são os elementos contidos num registro bibliográfico MARC 21, o que não parece tão simples”. (THOMALE, 2010, p. 3, tradução nossa).

O entusiasmo recém-desenvolvido para RDF se tornar a base para os dados bibliográficos de uma biblioteca tem ocorrido uma série de esforços e aplicações que convertam MARC 21 para RDF, mas nenhuma oficialmente ou recomendada pelos organismos internacionais da Catalogação Descritiva.

Dentre as iniciativas oficialmente aceitas da transformação dos dados em RDF encontram-se na *Library of Congress* (LC) padrões para a descrição de recursos, tais como o MODS (*Metadata Object Description Standard*). Outras propostas para a transformação dos dados de bibliotecas para RDF podem ser vistas em: ISBD² (*International Standard Bibliographic Description*) em RDF, FRBR³ em RDF e RDA⁴ em RDF.

Cada um desses esforços toma uma biblioteca padrão e usa RDF como sua tecnologia subjacente, criando um esquema de metadados completo que define cada elemento do padrão em RDF. O resultado é que se tem uma série de "silos" RDF, e cada elemento de dados definido como se pertencessem exclusivamente a esse padrão.

Existem quatro declarações diferentes, por exemplo, no elemento “lugar de publicação” em ISBD, RDA, FRBR e MODS, cada um com seu próprio URI e também há

¹ O termo dados ligados (*Linked Data*) refere-se ao conjunto das melhores práticas para a publicação e a interligação de dados estruturados na *Web*. Essas melhores práticas foram introduzidas por Tim Berners-Lee em sua nota *Web Architecture* e tornaram-se conhecidas como princípios de dados ligados. (HEATH; BIZER, 2011, tradução nossa).

² Maiores informações no site: Disponível em: <<http://www.ifla.org/en/node/1795>>. Acesso em: 20 jan. 2012.

³ Maiores informações no site: Disponível em: <<http://metadataregistry.org/schema/show/id/5.html>>. Acesso em: 20 jan. 2012.

⁴ Maiores detalhes no site: Disponível em: <<http://metadataregistry.org/schema/show/id/1.html>>. Acesso em: 20 jan. 2012.

diferenças entre elas (por exemplo, RDA separa lugar de publicação, produção etc., enquanto ISBD não o faz), certamente deverão ter uma estrutura (vocabulário) comum para solucionar essas questões. (COYLE, 2012).

Uma possível solução seria se as diferentes instâncias do "lugar de publicação" pudessem ser tratadas como tendo um significado comum, tal que elementos de FRBR pudessem ser ligados a um elemento ISBD, mas o mesmo não ocorre.

A razão pela qual não ocorre se justifica, porque cada uma dessas restringe os elementos de forma única (individual) que define seu relacionamento com um contexto de dados particular (o que geralmente pensamos das estruturas de um registro). Os elementos não são independentes do contexto, e isso significa que cada um somente pode ser usado dentro desse contexto particular. Essa é a antítese do conceito de dados ligados, onde conjuntos de dados de fontes diversas compartilham elementos de metadados. E esta reutiliza os elementos que cria o "link" em dados ligados (vinculados). Para conseguir tal processo, os elementos de metadados precisam ser irrestritos por um contexto particular. (COYLE, 2012).

A ligação pode também ser alcançada através dos relacionamentos verticais, similares aos termos (mais amplos e mais restritos) num tesouro. Essa opção é menos direta, mas torna possível a "mistura" de conjuntos de dados que têm diferentes níveis de granularidade. No caso do "lugar de publicação" da ISBD, poderiam ser definidos mais amplamente para os três elementos de RDA que tratam esse separadamente. Coyle (2012) afirma que infelizmente isso não é possível devido à forma que ISBD e RDA foram definidos em RDF.

Coyle (2012) reforça que o resultado é que agora tem-se uma série de "silos" RDF, expressões dos dados em RDF que faltam a capacidade de cruzamento dos dados ligados, porque eles são obrigados a especificar estruturas de dados e pouco se ganha em termos de dados ligados sob a óptica bibliográfica. Não somente os esquemas RDF são incompatíveis com outros, ninguém será ligado a dados bibliográficos de comunidades fora de bibliotecas que publicarem seus dados na *Web*. O que significa a não ligação dos dados à Amazon, à *Wikipedia*, para citações em documentos.

Coyle (2012) diz que devido ao estágio inicial em que estamos no desenvolvimento de dados ligados para o ambiente das bibliotecas, têm-se duas opções em relação ao uso do RDF.

- Definir "super-elementos" que estejam acima dos registros de formatos e que não estão vinculados pelas restrições dos registros definidos em RDF. Neste caso, haveria um "lugar de publicação" geral que seria um "super" elemento correspondente a todos os lugares de publicação nos vários registros e seriam subordinados a um conceito geral de "lugar" utilizado amplamente. Para implementar a ligação, cada elemento de registro seria extrapolado para os seus "super-elementos".
- Definir primeiro os elementos de dados fora de qualquer formato de registro específico, e então usá-los em esquemas de registros. Neste caso haveria apenas uma instância do "lugar de publicação" e que seria utilizado nos vários registros bibliográficos, sempre que um elemento é necessário. O intercâmbio desses registros seria possível com a ligação dos dados de seus elementos componentes, e iria interagir com outros dados bibliográficos na *Web* utilizando os elementos RDF definidos e seus relacionamentos.

Coyle (2012) conclui ainda que precisa-se criar dados, não somente registros, e que necessita-se criar primeiramente os dados, em seguida, registros com os mesmos para aplicações de acordo com a necessidade de cada ambiente informacional. Esses registros irão operar internamente nos sistemas de bibliotecas, enquanto os dados têm o potencial para fazer conexões no espaço de dados ligados. Um esforço maior tem que acontecer em descobrir e definir os elementos de nossos dados e olhar para os diversos dados que deseja unir, no vasto universo de informações.

As bibliotecas têm seu foco nos registros bibliográficos que compõem os acervos institucionais, geralmente um documento complexo que atua como um substituto do catálogo, tal como um livro ou uma gravação de música. RDF, nesse contexto, não menciona nada sobre os registros, apenas diz que existem dados que representam coisas (recursos) e os relacionamentos entre essas coisas. O que é muitas vezes confundido é que qualquer coisa pode ser uma coisa em RDF, assim, o livro, o autor, a página, a palavra na página, qualquer uma, ou todas essas poderiam ser as coisas em seu universo. (COYLE, 2012).

As discussões acerca do futuro das bibliotecas digitais e sua configuração para a adoção de tecnologias da *Web Semântica* manifestam-se da necessidade que os ambientes e sistemas de informação criem uma estrutura de dados que aproveitem as potencialidades do RDF, permitam os relacionamentos de forma explicitada e promovam a interoperabilidade.

Vale destacar que os relacionamentos bibliográficos sempre existiram no interior dos catálogos bibliográficos entre os dados de um registro bibliográfico, por meio das regras e esquemas de catalogação, porém não explicitados aos sujeitos institucionais para o uso e (re) uso das informações nos ambientes de manipulação de dados bibliográficos e catalográficos.

3 Modelagem estrutural de dados em ambientes informacionais digitais

Ao pensar na criação e no desenvolvimento de um modelo de dados, nos deparamos com as questões de granularidade e análise dos dados a serem catalogados. O que não é algo novo no domínio bibliográfico, onde se desenvolveu um modelo de dados implícito nas regras de descrição (AACR2) e nos formatos de intercâmbio de dados bibliográficos (MARC 21).

Atualmente com o desenvolvimento de novas regras e padrões para a modelagem dos ambientes informacionais, tais como FRBR, FRAD e RDA, reconhece-se que há uma tendência na estruturação e na definição dos dados a serem catalogados preparando-os para movê-los e torná-los compatíveis com a *Web Semântica*.

Nas camadas intangíveis dos dados (definida na representação e descrição dos recursos informacionais), há um aumento na estrutura e na granularidade dos dados. Yee (2009) aponta que mais estrutura e mais granularidade possibilitam apresentações mais sofisticadas aos usuários dos sistemas e aumentam a possibilidade de produção de dados interoperáveis.

Qualquer mudança ou mapeamento que foi contratado para criar dados interoperáveis produziria um menor denominador comum (os dados mais simples e menos granular), e uma vez interoperáveis, não seria possível sua recuperação na íntegra, devido sua perda. Dados com menos estrutura e menos granularidade poderiam ser mais fáceis e mais baratos para aplicá-los e ter o potencial mais simples para as comunidades envolvidas. (YEE, 2009, p. 59, tradução nossa).

Vamos tomar como exemplo um nome pessoal. Conforme as regras de catalogação (AACR2), demarca-se o sobrenome, do nome próprio, registrando primeiro o sobrenome, seguido por uma vírgula e depois o nome. Essa quantidade de granularidade pode representar muitas vezes um problema para o catalogador numa cultura desconhecida, que não domina necessariamente as regras. Mais granularidade pode ocasionar situações ambíguas para os sujeitos que estão coletando os dados. “Outro exemplo é quanto ao gênero do criador, onde o catalogador poderia se deparar com uma situação, que não saberia se o definiria como masculino ou feminino”. (YEE, 2009, p. 59, tradução nossa).

Yee (2009) comenta que se adicionarmos uma data de nascimento e de morte, seja qual for, as utilizamos juntas em um subcampo \$d sem qualquer codificação separada para indicar qual é o nascimento e qual é a morte (embora um ocasional "b" ou "d" nos dirá esse

tipo de informação). Poderíamos fornecer mais granularidade para datas, mas tornaria o formato MARC 21 muito mais complexo e difícil de aprender.

Na representação do campo 100 (autor pessoal), por exemplo, a forma autorizada para descrever o conteúdo é definida da seguinte maneira:

100 1#\$a Adams, Henry, \$d1838-1918.

Nesse caso, o subcampo \$d (Datas associadas ao nome NR⁵), 1838 corresponde à data de nascimento, enquanto 1918 à data de falecimento do autor.

De acordo com Yee (2009), granularidade e estrutura também podem ocorrer "tensão" uma com a outra. Mais granularidade pode conduzir a uma menor estrutura (ou mais complexidade para manter a estrutura junto com a granularidade). Na busca de maior granularidade de dados que se tem agora, (RDA tenta apoiar RDF na codificação XML), têm sido atomizados os dados para torná-los úteis aos computadores, mas isso não necessariamente tornará os dados mais compreensíveis aos humanos.

Para ser útil aos humanos, deve ser possível agrupar e organiza-los de forma significativa para a catalogação, a indexação e a apresentação dos mesmos. Os desenvolvedores do *Simple Knowledge Organization System* (SKOS) referem-se ao montante de informações não estruturadas (isto é, legível por humanos) na *Web*, rotulando *bits* de dados como relacionamentos semânticos dos registros em uma máquina acionável, de forma que não necessariamente fornece o tipo de estrutura necessária para tornar os dados legíveis por humanos e, portanto, úteis para as pessoas na *Web*. (YEE, 2009).

Para reforçar seu pensamento, Yee (2009, p. 59, tradução nossa) afirma que:

Quanto mais granular os dados, menos o catalogador pode construir ordem, sequenciamento e a ligação dos dados; a codificação deve ser cuidadosamente projetada para permitir a ordem, o sequenciamento e a ligação dos dados desejados, para que a catalogação, a indexação e a apresentação sejam possíveis, o que poderia ser chamado de uma codificação dos dados ainda mais complexa.

No que tange à estrutura de dados, Yee (2009) diz que existem vários significados atrelados ao termo, conforme pode ser observado.

- Estrutura é um objeto de um registro (estrutura de documento), por exemplo, *Elings* e *Waibel* referem-se a "campos de dados... também referidos como elementos... que estão organizados em um registro por uma estrutura de dados".
- Estrutura é a camada de comunicação, ao contrário da camada de apresentação ou designação de conteúdo.
- Estrutura é o registro, o campo e o subcampo.
- Estrutura é a ligação de *bits* de dados em conjunto, na forma de vários tipos de relacionamentos.
- Estrutura é a apresentação dos dados de maneira estruturada, ordenada e sequenciada para facilitar a compreensão humana.
- Estrutura de dados é a forma de armazenamento dos dados em um computador para que ele possa ser usado eficientemente (isto é, como os programas de computadores usam o termo).

Conforme as definições apresentadas por Yee (2009), Castro (2012, p. 168) ressalta que,

⁵ (NR) Não Repetido, significa que só pode existir apenas um subcampo com apenas um código associado ao nome do responsável pela obra.

Chamar-se-á de estrutura de dados, a camada intangível de instanciamento dos dados bibliográficos modelados para a representação e a descrição, tal como os formatos e/ou padrões de metadados, de forma a permitir a interoperabilidade dos ambientes informacionais digitais, por agentes humanos e não humanos, garantindo interfaces mais acessíveis aos usuários para posterior recuperação, uso e (re) uso dos recursos informacionais.

Quando se menciona estrutura de dados, no domínio bibliográfico, pensa-se no modelo conceitual de dados, estabelecido pelos Requisitos Funcionais para Registros Bibliográficos (FRBR).

FRBR faz uso de um modelo entidade-relacionamento, o qual consiste em dois principais conceitos: "coisas" e relacionamentos. FRBR define 10 categorias "coisas", as quais são denominadas entidades: Obra, Expressão, Manifestação, Item, Pessoa, Entidade coletiva, Conceito, Objeto, Evento e Lugar.

As entidades podem ser compreendidas, por exemplo, como uma obra, um texto, um livro etc. Os atributos correspondem às características dos dados relacionados à entidade e servem para diferenciar o conteúdo intelectual ou artístico. Os relacionamentos descrevem as ligações entre uma entidade e outra, na facilitação de manuseio do recurso informacional pelo usuário em um sistema. (MORENO, 2006).

Moreno (2006, p. 35, grifo do autor) apresenta as principais terminologias no modelo conceitual de dados FRBR:

Obra é uma entidade abstrata, uma criação intelectual ou artística distinta. A entidade **Expressão** é a realização intelectual ou artística específica que assume uma obra ao ser realizada, excluindo-se aí aspectos de alteração da forma física. Uma **Manifestação** é a materialização de uma expressão de uma obra, ou seja, seu suporte físico, que podem ser livros, periódicos, kits multimídia, filmes, etc., que é representada pelo **Item**, um único exemplar de uma manifestação. As duas últimas entidades refletem a forma física, são entidades concretas, enquanto as duas primeiras refletem o conteúdo intelectual ou artístico.

A contextualização e a exemplificação dos FRBR podem ser interpretadas da seguinte maneira. Tomemos como exemplo, um romance literário (Obra), onde se tem o texto original e o mesmo fora traduzido ou tivera alguma modificação, tal como uma edição ilustrada (Expressão); as formas em que a obra está disponível podem ser encontradas/visualizadas em um formato impresso (manual) ou num formato eletrônico/digital (Manifestação); e no momento em que a obra está disponibilizada na estante de uma biblioteca, ou seja, os exemplares relativos àquela obra são denominados **Itens** ("materialização" do recurso bibliográfico). (CASTRO, 2012).

Os FRBR constituem em uma "nova" ⁶ abordagem para a Catalogação Descritiva contemporaneamente nos seus moldes convencionais, no sentido de propiciar uma recuperação mais efetiva e intuitiva dos itens bibliográficos, agindo como um bibliotecário de referência, ou seja, relaciona todos os materiais atrelados ao termo da busca, trazendo-os de uma só vez em uma única interface. Por exemplo, se um determinado autor além de livros, possui outras manifestações, tais como discos, CDs e DVDs, o sistema/agente permitirá através do modelo de relacionamento FRBR, no momento da busca realizada pelo usuário, relacionar todas essas manifestações e recuperá-las uma única vez, apresentando-as aos usuários.

⁶ "Nova" no sentido de explicitar os relacionamentos entre registros bibliográficos, uma prática que já acontece, por exemplo, entre os elementos do formato de intercâmbio MARC 21, mas de forma não visível aos usuários dos catálogos *on-line*.

Os bibliotecários que se acostumarem com os Requisitos Funcionais para Registros Bibliográficos - FRBR - provavelmente não encontrarão muita dificuldade na transição do modelo conceitual para o modelo RDF. (CASTRO, 2012).

Riley (2010) aprofunda o estudo comparativo terminológico do RDF no contexto da Biblioteconomia e da Ciência da Informação, especificando:

- Assunto: em bibliotecas, o que um recurso informacional abrange em termos de conteúdo; em RDF – o que diz uma declaração sobre alguma coisa (recurso informacional).
- Vocabulário: em bibliotecas, implica num certo tipo de vocabulário controlado (termos autorizados, estruturas hierárquicas, termos relacionados etc.); em RDF – definições muito mais flexíveis (inclui definições formais de classes e de propriedades de um recurso informacional).
- Classe: em bibliotecas, um esquema de classificação (Classificação Decimal de Dewey – CDD; Classificação Decimal Universal – CDU etc.) indicando o tópico geral ou área do conhecimento abrangido pelo recurso informacional; em RDF - um tipo ou categoria pertencente a um objeto ou recurso informacional.
- “Schema”: XML *Schema* define um conjunto de elementos destinados a ser usados juntos; RDF *Schema* define classes e propriedades destinadas a ser usadas em qualquer lugar, sozinhas ou em combinação.

A dificuldade em qualquer exercício de modelagem de dados, sobretudo, no domínio bibliográfico, encontra-se em decidir o que tratar como entidade ou classe e o que tratar como um atributo ou propriedade. FRBR decidiu criar uma classe chamada expressão para tratar qualquer mudança no conteúdo de uma obra.

Os FRBR em comparação ao modelo de dados RDF, encontra-se em harmonia, as entidades do FRBR são registradas como classes, enquanto as relações são registradas como propriedades.

FRBR em RDF⁷ acrescenta apenas três classes. Duas delas (“*Endeavor*” e “Entidade Responsável”) são super conjuntos de classes FRBR. *Endeavor* é uma generalização que pode ser relacionada à obra, expressão, ou manifestação, ou seja, uma classe cujos membros são qualquer um dos produtos da atividade artística ou criativa. “Entidade Responsável” é um termo mais geral que pode se relacionar a uma entidade coletiva ou a uma pessoa. Essas classes especificam mais claramente informações sobre o conteúdo intelectual de um recurso, sem necessitar fornecer informações adicionais. “A terceira classe que é adicionada é o assunto. Todas as três incluem alguma instância do assunto em seus esquemas. FRBR trata claramente assunto como um relacionamento”. (DAVIS; NEWMAN, 2005; COYLE, 2012).

Com base na sustentação teórica apresentada nesse texto, até o presente momento, fica evidente que a condição dos ambientes informacionais digitais interoperar seus dados é a modelagem conceitual, esta definida e codificada por uma série de requisitos funcionais estabelecidos pelas arquiteturas de metadados, regras e esquemas de descrição bibliográfica e ontologias, que fornecerá ambientes melhor estruturados, na garantia de recuperação de informações mais efetiva aos usuários (humanos e não humanos).

4 Proposta de arquitetura para descrição bibliográfica semântica: caminhos para interoperabilidade

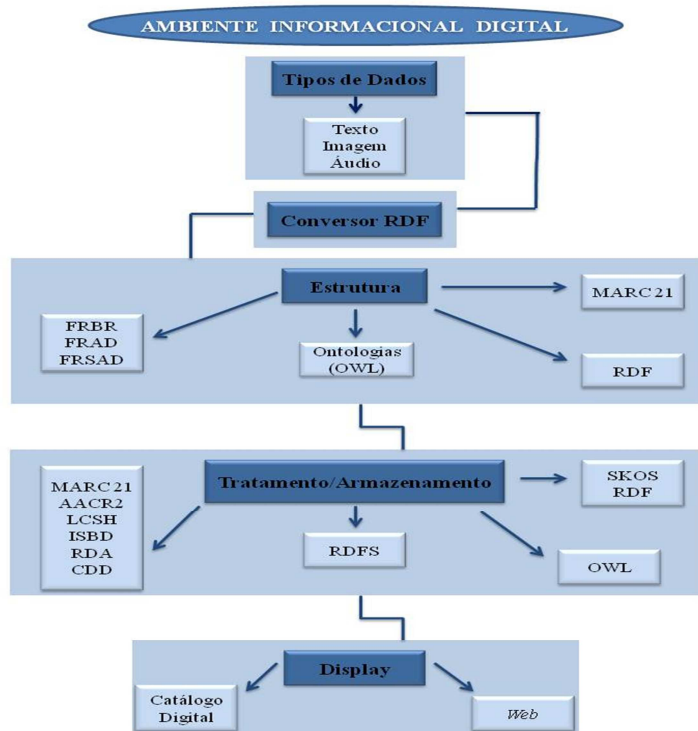
A palavra de ordem na constituição de ambientes informacionais digitais, sobretudo, no âmbito das bibliotecas digitais é a interoperabilidade. Para tanto, para que a

⁷ As expressões do núcleo de conceitos e vocabulários FRBR em RDF podem ser melhor visualizados no *site*. Disponível em: <<http://vocab.org/frbr/core.html>>. Acesso em: 20 jan. 2012.

interoperabilidade aconteça de maneira eficiente, faz-se necessário um olhar mais acurado nas camadas estruturantes e delineadoras dos catálogos bibliográficos digitais, ou seja, na representação e na descrição dos dados, a fim de potencializar as formas de busca e recuperação de informações.

Como um ponto de partida para garantir a interoperabilidade, esta investigação, a partir das observações e reflexões no campo da Catalogação Descritiva, destaca alguns requisitos e diretrizes funcionais que podem ser utilizados no estabelecimento da interoperabilidade em ambientes informacionais digitais de uma forma mais efetiva. Para tanto, pretende-se compreender a estrutura intangível, numa proposta apresentada em camadas sobrepostas, conforme a figura 1, uma vez que elas devem trabalhar em sinergia para a consistência e funcionamento pleno do ambiente digital.

FIGURA 1: Proposta de modelagem para interoperabilidade no domínio bibliográfico.



Fonte: Castro (2012, p.177)

- Camada 1 - Tipologia dos dados: nessa fase inicial o projetista (catalogador) define quais os dados serão trabalhados para alimentação e modelagem do ambiente informacional, a partir do recurso bibliográfico a ser catalogado. Exemplo: dados textuais, imagéticos, áudio etc. Vale destacar que nessa tese nos teremos apenas nos dados textuais, explicitados nos códigos de catalogação (AACR2) e nos padrões de metadados (MARC 21).
- Camada 2 - Preparação dos dados: uma vez definidos os dados bibliográficos a serem utilizados no sistema, a preparação dos dados consiste na adoção de ferramentas para a

conversão⁸ de dados em RDF. Os dados que foram extraídos de outras fontes, ou seja, dados não RDF, o W3C tem recomendado alguns conversores que auxiliam a transformação, como RDFizer. A adoção do RDFizer se justifica uma vez que não se tem oficialmente representações de dados MARC em RDF, nesse sentido, tal conversor transforma dados do padrão de metadados MODS em RDF, iniciativa essa promovida pela *Library of Congress* (MODS, 2011).

- Camada 3 – Tratamento e Armazenamento dos dados: Após a conversão dos dados em RDF, a próxima camada consiste na adoção efetiva pelo catalogador das regras e/ou dos esquemas de descrição bibliográfica (AACR2 e RDA), ou seja, a catalogação dos recursos bibliográficos, na confecção padronizada de metadados; a definição dos padrões de metadados (MARC 21), da arquitetura de metadados RDF para a estruturação dos dados e RDF *Schema* para sua validação. Vale dizer que nesse momento o catalogador deverá também adotar os Requisitos Funcionais para Registros Bibliográficos (FRBR), em consonância com os Requisitos Funcionais para Dados de Autoridade (FRAD) e os Requisitos Funcionais para Dados de Assunto (FRSAD). As ontologias aparecem nesse contexto para definir os conceitos dos elementos de um registro bibliográfico, por meio das regras e esquemas de descrição para a confecção metodológica dos metadados e dos padrões de metadados.
- Camada 4: Apresentação (*Display*) dos dados: A fase final consiste em tornar disponíveis os dados (*output*) e apresentá-los aos usuários do ambiente informacional. Os dados poderão aparecer da maneira como foram construídos e armazenados (*input*) nas camadas 1 e 3, para a camada tangível de recuperação e também visualizado na *Web*.

Acredita-se que esses requisitos e recomendações podem propiciar uma modelagem dos catálogos melhor estruturados, para posterior recuperação, uso e re (uso) das informações, garantindo a interoperabilidade e potencializando os relacionamentos bibliográficos semânticos, iniciativa essa que vai de encontro aos ideais vislumbrados pela *Web Semântica*.

5 À GUIA DE CONSIDERAÇÕES E REFLEXÕES

Acredita-se que tal modelo/esquema de arquitetura nessa investigação, pode possibilitar o compartilhamento entre padrões de metadados e ambientes e sistemas informacionais distintos, trabalhando numa filosofia de colaboração entre os recursos informacionais disponíveis e as tecnologias que estão abarcadas na sua construção, no estabelecimento da interoperabilidade, na otimização dos relacionamentos bibliográficos e ampliados para a construção padronizada de recursos na *Web*.

⁸ “Um conversor para RDF é uma ferramenta que converte aplicações de dados de um formato específico em RDF para uso com ferramentas de RDF e integração com outros dados. Conversores podem ser parte de um esforço de migração, ou parte de um sistema em execução que fornece uma visão *Web Semântica* de uma determinada aplicação”. (W3C, 2012, tradução nossa).

Segundo Castro (2012), modelos ontológicos de relacionamentos bibliográficos têm vantagens para a estruturação de novos catálogos. Uma ideia com muitas variações é estender campos do formato MARC 21 para ligações explícitas para outras obras. MARC 21 tem uma organização de arquivos de dados muito complexa. É possível converter a estrutura idiossincrática de registros MARC, ou seja, os códigos de campos, os subcampos e os valores condicionados aos indicadores para mais formatos padrão. A necessidade de fazê-lo, portanto, impede a utilização do avanço na corrente principal de tecnologias de banco de dados. (LEE; JACOB, 2011).

A utilização de ontologias, de metadados e dos fundamentos teóricos e metodológicos da Catalogação Descritiva uma vez articuladas, podem redesenhar novas opções aos ambientes informacionais, seja na modelagem do banco de dados (catálogos), ou ainda, na forma de representação dos recursos informacionais, garantindo a possibilidade de interoperabilidade semântica e no auxílio às formas de apresentação dos recursos, proporcionando aos usuários uma multiplicidade nas formas de busca, de acesso e de recuperação de informações pertinentes, significativas e relevantes, bem como o uso, a preservação e o (re) uso numa única interface. (CASTRO, 2012).

Quanto aos profissionais da Ciência da Informação, em especial os bibliotecários catalogadores, não podem prescindir de compreender a necessidade de construção de registros bibliográficos, que estejam pautados numa base teórico-epistemológica e metodológica que os sustentem e obedeça a registros pré-determinados e estabelecidos por regras e/ou esquemas de descrição bibliográfica internacionais, visando o controle bibliográfico universal.

De acordo com Castro (2012), a estrutura estratégica para promoção da interoperabilidade em ambientes informacionais digitais se dá na união da camada intangível (representação e descrição dos dados) de instanciamento e persistência dos dados bibliográficos, com a camada visível de apresentação para os usuários. O catalogador projeta por antecipação o ambiente informacional, utilizando ontologias para a descrição bibliográfica explicitadas nas regras e esquemas para a representação e descrição dos recursos bibliográficos, definindo conceitualmente os elementos, juntamente com os modelos conceituais e os padrões de metadados, garantindo a unicidade dos recursos, apresenta-os conforme as solicitações das estratégias de busca e de recuperação ao usuário.

Assim, a arquitetura para a representação e a descrição bibliográfica semântica e os níveis de interoperabilidade desenvolvida nesse estudo, propicia a modelagem estrutural dos ambientes informacionais digitais atualmente, a partir da heterogeneidade de esquemas de metadados, tornando-se, dessa forma, uma estrutura única a ser adotada por vários catalogadores, ampliando seu escopo para a construção padronizada de recursos na Web. Além disso, a adoção desse modelo configura o estabelecimento efetivo da interoperabilidade, principalmente com os impactos tecnológicos na Catalogação Descritiva.

REFERÊNCIAS

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, New York, May, 2001. Disponível em: <<http://www.sciam.com/2001/0501issue/0501berners-lee.html>>. Acesso em: 28 jun. 2005.

CASTRO, Fabiano F. de; SANTOS, Plácida L. V. A. C. Catalogação e metadados: interlocuções nos ambientes informacionais digitais. In: ARELLANO, Filiberto Felipe Martínez (Org.). *Memoria del IV encuentro de catalogación y metadatos*. México: Centro Universitario de Investigaciones Bibliotecológicas de la Univ. Nacional Autónoma de México, 2010. v. 4, p. 301-318.

CASTRO, Fabiano F. de. Elementos de interoperabilidade na catalogação descritiva: configurações contemporâneas para a modelagem de ambientes informacionais digitais. 2012. 202 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012.

COYLE, Karen. MARC 21 as data: a start. *Code4Lib Journal*, Vol. 14, 2011.

COYLE, Karen. *Bibliographic framework: RDF and linked data*. Berkeley, United States, Jan. 2012. Disponível em: < <http://kcoyle.blogspot.com/2012/01/bibliographic-framework-rdf-and-linked.html>>. Acesso em: 20 jan. 2012.

DAVIS, Ian; NEWMAN, Richard. *Expression of core FRBR concepts in RDF*. 2005. Disponível em: < <http://vocab.org/frbr/core.html>>. Acesso em: 20 jan. 2012.

HEATH, Tom; BIZER, Christian. *Linked data: evolving the web into a global data space*. Morgan & Claypool, Vol. 1 no. 1, 2011. 136 p. (Synthesis Lectures on the Semantic Web: Theory and Technology). Disponível em: < <http://linkeddatabook.com/editions/1.0/#linkedData>>. Acesso em: 15 nov. 2011.

LEE, Seungmin; JACOB, Elin K. An integrated approach to metadata interoperability: construction of a conceptual structure between MARC and FRBR. *Library Resources & Technical Services*, Vol. 55, no. 1, p.17-32, Jan. 2011.

MODS. *Metadata Object Description Schema*. Library of Congress, Mar. 2011. Disponível em: < <http://www.loc.gov/standards/mads/rdf/>>. Acesso em: 21 maio 2011.

MORENO, F. P. *Requisitos funcionais para registros bibliográficos - FRBR: um estudo no catálogo da rede bibliodata*. 2006. 202 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2006.

RILEY, Jenn. Enhancing interoperability of FRBR-based metadata. In: *Proc. Int'l Conf. on Dublin Core and Metadata Applications*. 2010a. p. 31-43. Disponível em: <<http://dcpapers.dublincore.org/index.php/pubs/article/view/1037>>. Acesso em: 12 dez. 2011.

THOMALE, Jason. *Interpreting MARC: where's the bibliographic data?* *Code4Lib Journal*, Vol. 11, 2010.

W3C. *Converter to RDF*. 2012. Disponível em: < <http://www.w3.org/wiki/ConverterToRdf>>. Acesso em: 21 jan. 2012.

YEE, Martha M. Can bibliographic data be put directly onto the semantic web? *Information Technology and Libraries*, Vol. 28, no. 2, p. 55-80, Jun. 2009.