



Revisitando ontologia e metadados à luz dos ambientes informativos digitais

Fabiano Ferreira de Castro^I

<http://orcid.org/0000-0002-8712-2654>

Ana Carolina Simionato Arakaki^{II}

<http://orcid.org/0000-0002-0140-9110>

^I Universidade Federal de São Carlos, SP, Brasil.

Doutor em Ciência da Informação. Professor Adjunto do Departamento de Ciência da Informação e do Programa de Pós-Graduação em Ciência da Informação (PPGCI/UFSCar).

^{II} Universidade Federal de São Carlos, SP, Brasil.

Doutora em Ciência da Informação. Professora do Departamento de Ciência da Informação e do Programa de Pós-Graduação em Ciência da Informação (PPGCI/UFSCar).

<http://dx.doi.org/10.1590/1981-5344/3329>

O cenário contemporâneo marcado pela explosão informacional torna-se imprescindível que os ambientes informativos altamente distribuídos, como a Web, compartilhem e colaborem com dados e informações organizadas e estruturadas. Tal realidade tem desafiado profissionais de diversas áreas do conhecimento, sobretudo, os da área da Ciência da Informação a buscar soluções para o tratamento informacional de tais ambientes, a fim de garantir mecanismos para a busca e a recuperação de recursos informativos eficientes. Elucida-se que 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

informativos. Objetiva-se verificar a dimensão relacional entre os instrumentos para a representação de dados e informações, tais como a ontologia e os metadados, considerados como pilares para a modelagem, a estruturação de catálogos e sistemas bibliográficos. Por meio de uma metodologia exploratória e descritiva interdisciplinar nas áreas da Ciência da Informação e da Ciência da Computação identificam-se elementos conceituais, a fim de apontar a necessidade da construção de ambientes informativos digitais estruturados e padronizados. A utilização da ontologia e dos metadados de forma sinérgica apresenta importância capital para a representação e a descrição de recursos bibliográficos nos ambientes informativos digitais atualmente, proporcionando a modelagem dos catálogos bibliográficos e a interoperabilidade semântica entre diferentes sistemas e plataformas.

Palavras-chave: *Ontologia. Metadados. Modelagem de ambientes informativos digitais. Interoperabilidade. Informação e Tecnologia.*

Revisiting ontology and metadata in the light of digital informational environments

The contemporary scenario marked by the informational explosion it is essential that highly distributed information environments such as the Web, share and collaborate with organized and structured data and information. This reality has challenged professionals from different fields of knowledge, especially the area of Information Science to seek solutions to the informational treatment of such environments in order to ensure efficient search and retrieval of informational resources. Elucidated that the treatment of the information resource requires a description of form and content readable by machines with human understandable results, and it also meets the requirements of interoperability between informational

environments. The objective is to check the relational dimension between the instruments for the representation of data and information, such as ontology and metadata, regarded as pillars for modeling, structuring catalogs and bibliographic systems. Through an exploratory and descriptive methodology interdisciplinary in the areas of Information Science and Computer Science are identified conceptual elements in order to point out the need for the construction of structured digital information environments and standardized. The use of ontology and metadata synergistically presents vital importance for the representation and the description of bibliographic resources in digital informational environments currently, providing a modeling of bibliographic catalogs and semantic interoperability between different systems and platforms.

Keywords: *Ontology. Metadata. Modeling digital information environment. Interoperability. Information and Technology.*

Recebido em 31.07.2017 Aceito em 05.10.2020

1 Introdução

O tecido social contemporaneamente é marcado pelo uso intensivo das Tecnologias da Informação e Comunicação (TIC) e geram, por consequência, o desenvolvimento de ambientes informativos digitais com mais recursos e ferramentas para a produção e o uso da informação.

Nesse cenário, percebe-se que houve um aumento desordenado e caótico na quantidade de informações produzidas e disponibilizadas em meio digital, requerendo uma mudança e um repensar nas formas de armazenamento, de representação, de descrição e de preservação e que garantam o acesso aos recursos informativos digitais.

A evolução das bibliotecas tem merecido grande destaque, ao longo do século XX, no que se refere ao desenvolvimento e uso de tecnologias, principalmente as TIC, pois potencializaram seus serviços com o oferecimento de novos recursos de acesso, com a utilização de formatos de intercâmbio de dados na análise e na representação da informação, no

uso de ontologias, na orientação para a modelagem de catálogos e no processo de busca e de recuperação da informação.

Dessa forma, há uma preocupação da comunidade científica, no que tange à busca por caminhos e soluções para as questões de integração, interoperabilidade e entendimento semântico sobre os recursos informativos disponíveis hoje nos ambientes informativos digitais, de forma a possibilitar o acesso, o uso e o reuso das informações de forma mais efetiva para o usuário final.

Elucida-se que o tratamento dado ao recurso informativo 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 informativos.

Considerando áreas do conhecimento científico que se preocupam com os aspectos intelectuais da descrição de dados e da informação para posterior recuperação e caracterizadas pela interdisciplinaridade, Saracevic (1995) aponta que existe um componente significativo na Ciência da Informação e na Ciência da Computação, associado com a representação da informação e no uso otimizado dos instrumentos e das ferramentas que garantem a padronização do recurso informativo, bem como sua organização intelectual e suas relações no ambiente digital.

Diante do exposto, essa pesquisa teve como objetivo verificar a dimensão relacional entre a ontologia e os metadados, considerados como pilares para a modelagem e a estruturação de catálogos e sistemas bibliográficos, contemporaneamente.

Caracterizada por uma metodologia de análise exploratória e descritiva foi possível identificar elementos conceituais, a partir da literatura científica das áreas de Ciência da Informação e da Ciência da Computação, a fim de apontar a necessidade da construção de ambientes informativos digitais estruturados e padronizados, na utilização de ontologia e metadados.

Destaca-se que a ontologia e os metadados apresentam-se como importantes instrumentos para a representação e a descrição de dados e de informações nos ambientes informativos digitais atualmente, proporcionando, se utilizados em conjunto, a interoperabilidade semântica entre diferentes sistemas e plataformas.

2 Ontologia: requisito para estruturação de ambientes digitais

No contexto da integração de dados, a heterogeneidade é fator essencial a ser tratado, tendo em vista que as fontes a serem integradas foram desenvolvidas em contextos distintos, por pessoas e organizações diferentes. Assim, são encontrados modelos de dados diversos, representações diferentes para os dados, terminologias conflitantes etc.

A integração de dados, principalmente em ambientes informativos digitais, tem sido objeto de investigação e pesquisa, dinâmica e desafiadora durante muitos anos, com o intuito de fornecer aos usuários interfaces uniformes e mais amigáveis de acesso, relacionando e combinando os diferentes recursos.

As fontes de dados podem ser heterogêneas na sintaxe, no esquema ou na semântica, tornando uma tarefa difícil à interoperabilidade dos dados. A heterogeneidade sintática é causada pelo uso de diferentes modelos ou linguagens. A heterogeneidade esquemática é resultado das diferenças estruturais. Enquanto isso, a heterogeneidade semântica é causada pelos diferentes significados ou interpretação dos dados em vários contextos. Para alcançar a interoperabilidade dos dados é necessária a eliminação de sua heterogeneidade. (CRUZ; XIÃO, 2005, tradução nossa).

Nesse sentido, e observando as questões de interoperabilidade semântica entre ambientes informativos digitais, destaca-se que dentre as heterogeneidades, a que representa um obstáculo maior para a integração de esquemas e, conseqüentemente, para a integração de dados, é a heterogeneidade semântica, uma vez que esta é marcada por conflitos semânticos que podem ocorrer quando dois contextos não usam a mesma interpretação da informação.

A integração semântica de dados seria “[...] o processo de utilização de uma representação conceitual dos dados e seus relacionamentos, com o objetivo de eliminar as possíveis heterogeneidades de compreensão e interpretação”. (CRUZ; XIÃO, 2005, p. 02, tradução nossa). Ela ainda pode ser entendida como aquela que inclui a semântica das requisições de informação do usuário e das fontes de informação.

Chateaubriand (1998) afirma que uma das tendências do século XXI é fazer a ontologia, derivar da integração semântica, uma vez que há uma relação entre a linguagem e a realidade. Como forma de possibilitar a adição de semântica aos metadados e a integração de esquemas, tem sido proposto o uso de ontologia, em decorrência do seu potencial para descrever a semântica de fontes de informação distintas e como forma de resolver os problemas de heterogeneidade na *Web*. (CRUZ; XIÃO, 2005).

No centro da integração semântica dos dados está o conceito de ontologia, ou seja, uma especificação explícita de uma conceitualização compartilhada, com o propósito de facilitar o compartilhamento do conhecimento em áreas distintas para o reuso dos dados.

Uma ontologia define os termos usados para descrever e representar uma área do conhecimento. Ontologias são usadas por pessoas, bancos de dados e aplicações que necessitam compartilhar informações de domínio (um domínio é apenas um assunto específico de uma área ou

área de conhecimento, como a medicina, gestão financeira etc.). Ontologias incluem definições de conceitos básicos processáveis pelo computador no domínio e os relacionamentos entre eles. As ontologias codificam o conhecimento em um domínio e também o conhecimento que cobrem os domínios. Dessa forma, faz com que o conhecimento seja reutilizável. (OWL, 2004, tradução nossa).

Existe uma pluralidade semântica acerca do conceito de ontologia, de acordo com a área do conhecimento especializado. Na Ciência da Computação, o termo representa um artefato de engenharia de *software* (GRUBER, 1993). Para Gruber (1993), ontologia é uma especificação de uma conceitualização, ou seja, é uma descrição de conceitos e relacionamentos que existem entre estes conceitos.

A palavra-chave na para fins de construção de ontologia é compartilhamento e reutilização do conhecimento. Para permitir compartilhamento e reutilização de conhecimento, em Ciência da Computação, é importante que os conceitos, presentes nas ontologias possuam de fato uma especificação formal.

Para tanto, Gruber (1993, p. 202, tradução nossa) sugere uma maneira de criar definições formais para conceitos.

Seja a estrutura tal que D representa o domínio em questão e W representa todos os conceitos existentes em D. Esta estrutura é chamada de Espaço de Domínios. Uma conceitualização C é uma estrutura sendo R o conjunto de relações escolhidas como pertinentes para representar este domínio. Desta forma, uma conceitualização define uma estrutura pretendida do mundo, representada por S. Do ponto de vista computacional, para que uma conceitualização possa ser efetivamente usada, ela precisa ser especificada em uma determinada linguagem L. A estrutura de S é, então, mapeada para constantes e predicados da linguagem L, seguindo uma função de interpretação. (GRUBER, 1993, p. 202).

Uma das principais infraestruturas da *Web Semântica*, em relação às questões da integração semântica são as ontologias que se apresentam como instrumentos para a descrição de recursos informativos. Segundo Cunha (2002), uma ontologia fornece um entendimento comum e compartilhado de um domínio, que pode ser comunicado através de pessoas e sistemas de aplicação, tornando-se elemento-chave para o desenvolvimento *da Web Semântica*.

A ontologia no campo da Ciência da Informação, tem sido estudada desde o século XIX, principalmente em estudos bibliográficos para a representação dos sujeitos (VICKERY, 1997; ALMEIDA 2013, 2014) e

oferece soluções para o problema da heterogeneidade semântica, que leva em consideração o significado dos dados armazenados e sua interpretação para os usuários humanos e não humanos.

Na ciência da informação, os princípios ontológicos podem ser utilizados para apoiar a construção de estruturas categóricas para a representação do conteúdo dos documentos. O trabalho ontológico em ciência da informação não se limita a representação do conteúdo temático de recursos, mas engloba representação do recurso como um todo, em geral, a partir da perspectiva de uma determinada comunidade de usuários. Ontologia, nesse sentido, é um sujeito autêntico e fecundo de pesquisa em ciência da informação, uma vez que tem o potencial de explicar não só as questões relacionadas com o conteúdo de um documento, mas todo o ambiente social envolvido na análise conduzida por cientistas da informação. (ALMEIDA 2013, p. 1691, tradução nossa).

A partir de Almeida (2013, 2014) defende-se que a ontologia vai além da representação de conteúdo dos recursos informativos, pois irá atuar nas camadas de descrição do recurso, que sustenta o delineamento e a construção da ontologia, ou seja, na modelagem conceitual dos atributos e dos dados que garantem a persistência e a padronização dos elementos descritivos nos catálogos e em sistemas bibliográficos, na definição dos relacionamentos e na interoperabilidade em ambientes informativos.

Dessa forma, as ontologias podem ser utilizadas na integração de arquiteturas de metadados, como um esquema global para o mapeamento de diferentes fontes de metadados. Em comparação a outros esquemas, a ontologia conceitua os domínios de interesses particulares e semanticamente ricos. (CRUZ; XIÃO, 2005).

De acordo com Corcho, Fernández-Lopez e Gómez-Pérez (2001, tradução nossa), uma ontologia é constituída de 5 (cinco) componentes centrais, dentre os quais se destacam: classes, relacionamentos, funções, axiomas e instâncias e podem ser entendidas no domínio da representação da informação, de acordo com o Quadro 1.

Quadro 1 - Analogia da ontologia com a representação da informação.

Classes	São comumente organizadas em taxonomias. As classes ou conceitos são utilizados num sentido mais geral, ou seja, podendo ser o conceito propriamente dito sobre alguma coisa ou a descrição de tarefas, funções, ações, estratégias, processo de raciocínio etc. Um exemplo, pode ser a
----------------	---

	tipologia dos recursos informativos, como livros, periódicos, mapas, audiovisuais, entre outros.
Relações	Representam o tipo de envolvimento ou interação entre os conceitos de um domínio. Elas podem ser formalmente definidas entre subconjuntos, tal como, as relações do tipo ‘ <u>subclasse de</u> ’ ou ‘ <u>conectada a</u> ’. No domínio bibliográfico podem-se ver as relações entre os elementos descritivos de um padrão de metadados, por exemplo, no <i>MachineReadableCatalogin</i> (MARC 21), obra e seus assuntos.
Funções	Apresentam-se como um caso especial de relação onde o enésimo elemento da relação é único para os n-1 elementos precedentes. Exemplo: ‘antecedente-de’ e causa. Uma obra pode ter um ou mais autores, um autor pode ter uma ou mais obras.
Axiomas	São utilizadas para modelar as sentenças que são sempre verdadeiras e aplicadas de acordo com várias propostas: definir exatamente o significado dos componentes de uma ontologia, as restrições complexas nos valores dos atributos etc.; de forma a verificar a correção das informações especificadas na ontologia, ou ainda, deduzindo novas informações. Exemplo: As regras e/ou esquemas de descrição bibliográfica, <i>Anglo-AmericanCataloguingRules</i> (AACR2r), <i>ResourceDescription & Access</i> (RDA), etc. definem os valores dos atributos ou dos elementos de metadados.
Instâncias	São empregadas para representar elementos específicos, ou seja, os dados propriamente ditos. Na representação bibliográfica têm-se os metadados tais como, autor, título, ano de publicação, assunto, etc.

Fonte: Adaptado de Corcho, Fernández-Lopez e Gómez-Pérez (2001).

Considera-se nesse trabalho, a ontologia como um artefato tecnológico que propiciará a modelagem de ambientes informativos digitais, através do estabelecimento conceitual na definição dos elementos de metadados, no processo da descrição dos dados e das informações, para uma compreensão/leitura da semântica para usuários humanos e não humanos, oportunizando a interoperabilidade em ambientes informativos digitais.

Para Souza e Alvarenga (2004) a ontologia se apresenta como um modelo de relacionamento de entidades e suas interações, em algum domínio particular do conhecimento ou específico a alguma atividade. O

objetivo de sua construção é a necessidade de um vocabulário compartilhado para trocarem informações entre os membros de uma comunidade, sejam eles seres humanos ou agentes inteligentes.

No domínio bibliográfico, por exemplo, uma ontologia atuaria na definição dos conceitos dos atributos de um padrão de metadados de conteúdo, refletindo principalmente no valor de seus dados, tal como as etiquetas do formato de intercâmbio MARC 21, ou seja, o vocabulário poderia ser definido de acordo com as regras e/ou esquemas de descrição contempladas pelo AACR2r ou RDA.

Uma ontologia tem por objetivo prover uma base semântica para esquemas de metadados, facilitando a comunicação entre sistemas e agentes. Assim, a ontologia estabelece fundamentos de significados conceituais sem os quais a *Web Semântica* não seria possível, devido à heterogeneidade dos conceitos representados. (JACOB, 2003, tradução nossa).

Uma ontologia tem diferentes níveis de estrutura. Algumas ontologias são taxonomias, isto é, classificações ou listas hierárquicas de terminologias de uma área particular. Algumas são especificações de conjuntos de características conceituais. "Por exemplo, os esquemas de metadados são ontologias que especificam elementos para ser usados, o que estes elementos significam, e quais tipos de atributos e valores aqueles elementos podem ter". (TAYLOR, 2004, p. 284, tradução nossa). Outras ontologias aparecem para categorizar vocabulários controlados.

Quando se exemplifica a ontologia como esquemas de metadados (TAYLOR, 2004), começa-se a perceber as sutis aproximações e suas correspondências, sobretudo, nas questões de representação de recursos e, ainda, os possíveis relacionamentos que só se efetivarão em um trabalho conjunto; pode-se dizer que há uma sinergia entre ambas, principalmente no delineamento e na modelagem de ambientes informativos digitais e que poderão proporcionar a interoperabilidade semântica.

É sob essa orientação que se percebe a ligação entre ontologia e metadados, uma vez que para atender aos requisitos de interoperabilidade semântica entre padrões e/ou esquemas de metadados heterogêneos, é imprescindível uma base ontológica que suporte e que defina as variadas terminologias, como os campos, subcampos, indicadores e outros elementos de um registro bibliográfico, indispensáveis para uma representação e apresentação adequadas.

De acordo com as observações apresentadas sobre ontologia e o seu papel na interoperabilidade semântica, pode-se dizer que essa define as entidades em um nível abstrato, com a intenção de conceituar um domínio de interesse particular. Ela não fornece elementos específicos para a

descrição de um recurso, mas uma visão geral de noções básicas de um campo de interesse e as relações entre essas noções. (CASTRO, 2012).

Reforçando a correlação entre ontologia e metadados, Greenberg, Sutton e Campbell (2003) apontam ainda que a ontologia é compreendida como sistemas de metadados ou vocabulários de metadados, constituindo na visão dos autores “[...] a artéria central de metadados para a *Web Semântica*”.

A ontologia permite uma melhor representação para os metadados de sistemas de integração: maior semântica, especialmente no que diz respeito às fontes e aos mapeamentos e, na integração de esquemas de metadados (*metadataschema*). Enquanto os metadados são usados para descrever e representar recursos em termos de elementos, e facilitar a descoberta e o fácil acesso à informação.

3 Metadados: exigência para a representação da informação na era digital

Os bibliotecários produzem e padronizam metadados há séculos, desde as primeiras tentativas de organização da informação a partir da descrição de documentos. O que vem acontecendo ultimamente é que profissionais de diversas áreas estão buscando criar instrumentos de descrição da informação, mas seu desconhecimento dos métodos, processos e peculiaridades característicos da documentação, da Biblioteconomia, tem gerado uma variedade de padrões que muitas vezes não atende satisfatoriamente às exigências de uma lógica descritiva estabelecida e que dê conta da complexidade da caracterização desse material. (MILSTEAD; FELDMAN, 1999; ALVES, 2005; CASTRO; SANTOS, 2007; CASTRO, 2008).

Pode-se dizer que o objetivo e a função dos metadados estão fundamentados nos princípios da catalogação, ou seja, garantir a padronização dos recursos informativos (forma e conteúdo), pautados em regras, códigos e esquemas internacionais na tentativa de facilitar e potencializar a identificação, a busca, a localização, a recuperação, a preservação, o uso e o reuso dos recursos informativos. A diferença dessa forma de representação está na nova abordagem dada pelo ambiente tecnológico em que ela se insere. (ALVES, 2005; CASTRO, 2008).

Metadados não podem ser descrições desestruturadas de recursos, eles devem ser padronizados e controlados. Para Senso e Rosa Piñeros (2003, p. 97, tradução nossa) os metadados são como “[...] toda a informação descritiva sobre o contexto, qualidade, condição ou características de um recurso, dado ou objeto”. Os metadados podem descrever uma variedade imensa e diversificada de informações, como o assunto do recurso informativo, os criadores do recurso, a informação

técnica necessária para armazenar e acessar o recurso, os direitos legais do recurso. (LUBAS; JACKSON; SCHNEIDER, 2013).

As unidades básicas dos metadados são os esquemas e os elementos. Elementos de metadados são as categorias individuais ou classes que mantêm ou sustentam as partes individuais da descrição de um conjunto/pacote de informação. Elementos de metadados típicos incluem, por exemplo, título, criador, data de criação, identificação de assunto e a preferência. Na visão de Zeng e Qin (2008, p. 321, tradução nossa) o elemento de metadado (*metadatelement*) “[...] é um termo formalmente definido para descrever uma das propriedades de um recurso de um tipo particular ou para uma proposta particular. Por exemplo, o „publicador“ de um livro, „o formato“ de um arquivo eletrônico, ou uma „restauração de dados“ de uma construção”.

Enquanto isso, os esquemas de metadados são conjuntos de elementos desenhados para encontrar as necessidades de comunidades particulares. Enquanto alguns esquemas são gerais em sua natureza, têm sido criados mais esquemas para tipos de informação específica. “Esquemas têm sido desenhados para lidar com informação governamental, geoespacial, busca visual e muitos outros tipos de pacotes/conjuntos de informações”. (TAYLOR, 2004, p. 147, tradução nossa). Esquemas podem variar grandemente, por exemplo, no número de elementos de dados, no uso de elementos obrigatórios e repetíveis, em codificação e no uso de vocabulários controlados, entre outras coisas. Enquanto mais esquemas focam em elementos descritivos para suportar/sustentar a descoberta do recurso, alguns contêm elementos para suportar propostas administrativas e estruturais. “Com as várias necessidades de diferentes comunidades, não é possível criar um esquema perfeito, que contempla a totalidade do metadado”. (TAYLOR, 2004, p. 145, tradução nossa).

De acordo com Zeng e Qin (2008, p. 323, tradução nossa) o esquema de metadados pode ser considerado como:

Uma especificação processável por máquinas que define a estrutura, a codificação de sintaxe, regras, e formatos para um conjunto de elementos de metadados em uma linguagem formal num esquema. Na literatura o termo esquema de metadados usualmente refere-se ao conjunto de elementos na sua totalidade, assim como a codificação dos elementos e a estrutura com uma linguagem de marcação. (ZENG; QIN, 2008, p.323).

De acordo com Vellucci (2000, tradução nossa) há três características encontradas em todos os esquemas de metadados: (1) estrutura (2) sintaxe (3) semântica.

(1) Estrutura: refere-se ao modelo de dados ou arquitetura usada para manter/comportar o metadado e a forma como as declarações (*statements*) do metadado são expressas. Estrutura aqui é entendida ou está sendo referida à estrutura do metadado. Não deve ser confundida com „metadado estrutural“, que se refere à estrutura descrita inicialmente do recurso. Dois exemplos de tais modelos são RDF e *Metadata Encoding and Transmission Standard* (METS).

(2) Sintaxe: refere-se à codificação do metadado. Este pode ser o formato MARC 21 para registros bibliográficos ou um *extensible Markup Language* (XML) ou *Standard Generalized Markup Language Document Type Definition* (SGML DTD)¹ para outros tipos de metadados.

(3) Semântica: refere-se ao significado, especialmente o significado de vários elementos de dados. A semântica ajuda aos criadores a entender, por exemplo, o que „cobertura“ ou „modificação de data“ significa em um dado esquema. A semântica de um esquema de metadados não dita o espaço do conteúdo junto aos elementos. Esta é a „província“ de padrões de conteúdo (ou regras de conteúdo) e vocabulários controlados. Os padrões de conteúdo determinam tais coisas, assim como a data será formatada junto aos elementos do metadado. Por exemplo, ele deve especificar que todas as datas serão entradas usando o formato YYYY-MM-DD, sendo YYYY o ano; MM o mês; e DD o dia. Vocabulários controlados se referem a listas de palavras os quais certos termos são escolhidos e seus sinônimos, limitando a extensão de valores que podem ser inseridos dentro de uma classe. Vocabulários controlados são frequentemente usados em tipos de objetos e assuntos relacionais de elementos de dados. Se tais regras e sistemas não existissem, a recuperação da informação de uma forma efetiva poderia ser comprometida.

Para exemplificar as características encontradas num esquema de metadados, tem-se a figura 1, com o atributo (245) do formato de intercâmbio bibliográfico MARC 21. O campo, etiqueta, metadado ou atributo² (245) equivale ao título do recurso bibliográfico e possui os subcampos (a) informação sobre o título do recurso; (b) subtítulo do recurso; e (c) indicação de responsabilidade sobre o recurso bibliográfico. Os campos e subcampos no esquema MARC 21 correspondem à sintaxe do registro bibliográfico, e precisam ser preenchidos de acordo com as regras de conteúdo (AACR2r) que definem e padronizam a descrição e a representação bibliográfica (semântica).

¹ *Document Type Definition* (DT D) estabelece um conjunto de regras (sintaxe) para a validação da estrutura de um documento XML. Uma das desvantagens da DT D é que ela permite apenas a verificação sintática de um documento sem nenhum tipo de controle semântico (conteúdo). “[...] pode-se garantir que o título foi digitado, mas não se realmente era o título, ou ainda, pode-se verificar se o dado informado no campo „data de publicação“ é realmente uma data, mas não se garante que seja a data correta”. (SIQUEIRA, 2003, p. 75).

² Principais terminologias encontradas na literatura de Catalogação e Metadados para caracterizar a constituição de um padrão e/ou esquema de metadados.

Figura 1 - Etiqueta título do formato MARC 21.

245 00 |a Introduction to metadata: |b pathways to digital information / |c edited by

Fonte: Elaborado pelos autores.

De acordo com as características dos esquemas de metadados, Zeng e Qin (2008, p. 93, tradução nossa) apontam que dentre elas, duas não podem ser esquecidas, consideradas componentes básicos no conjunto de elementos do esquema, as quais se destacam: “[...] Semântica (definições dos significados dos elementos e seus relacionamentos); e Conteúdo (declarações ou instruções do que e como os valores devem ser atribuídos aos elementos)”.

No que concerne à semântica, Zeng e Qin (2008) apontam que os esquemas de metadados são criados para a identificação e a descrição de recursos e não expressam semântica rica³. Mesmo que o significado das informações dos metadados e seus relacionamentos para descrever recursos podem ser entendidos e processados por humanos, o processamento por máquinas e seus relacionamentos não é óbvio, ao menos que declarado. Ao contrário, as classes de ontologia são interligadas por propriedades específicas que declaram explicitamente os relacionamentos semânticos entre essas entidades. Mapeando um esquema de metadado real para uma ontologia gera muitas dificuldades devido ao grande número de expressões conceituais que devem ser alinhados.

De acordo com Castro (2012), bons modelos de estruturas é resultado do uso correto de princípios, os quais podem ser encontrados e há muito tempo consolidados nas metodologias cristalizadas e sedimentadas das áreas de Biblioteconomia e Ciência da Informação, por meio do uso dos códigos de catalogação, por exemplo, os quais padronizam a elaboração e a construção dos metadados e dos esquemas de metadados nos variados ambientes informativos (digitais ou não).

Com a heterogeneidade dos recursos informativos, armazenados e estruturados nos mais diversos esquemas, a literatura científica aponta algumas técnicas de mapeamento de elementos de metadados, entre padrões de metadados distintos. Dentre elas destacam-se especialmente, o *crosswalk*, pois é um método que contribui para os aspectos de

³A expressão rica (o) refere-se às estruturas com maior número de elementos de descrição e seus dados são codificados por algum esquema de descrição bibliográfica. Exemplo: Estrutura MARC 21 e esquema AACR2r.

interoperabilidade em ambientes informativos digitais e entre os esquemas de metadados.

A *National Information Standards Organization* (NISO, 2004) define um *crosswalk* como um “[...] mapeamento de elementos, semântica e sintaxe de um esquema de metadados para outros”. Sendo assim, o método *crosswalk* permite “[...] a identificação formal dos elementos de metadados ou grupos de elementos destes de diferentes esquemas equivalentes para que sejam semânticos [...]” (SIMIONATO, 2015, p. 111) e para que abranja cada contexto.

O método inicia pelo estudo dos esquemas de metadados de cada domínio; e a partir disso, é necessário comparar os elementos descritivos ou elementos fixos a um ou mais padrões, deixando as diferenças e as semelhanças entendidas em vários níveis, de forma a avaliar o grau em que os sistemas são interoperáveis e representar visualmente essas relações de semelhanças e diferenças. (BACA, 2008; SIMIONATO, 2015).

A tabulação dos elementos descritivos é o mecanismo mais comum para o *crosswalk*, de forma que os resultados representam um mapeamento semântico dos elementos de dados em outro padrão de dados (fonte/origem) para aqueles em outro padrão (alvo) baseado na similaridade da função ou significado dos elementos. (BACA, 2008; ZENG; QIN, 2008).

Um exemplo de *crosswalk* está na figura 2, a correspondência entre o campo título para os esquemas da ISBD Consolidada, *Grafic Materials*, AACR2r e RDA.

Figura 2 - Correspondência para o título

ISBD Consolidada	Grafic Materials	AACR2r	RDA
1.1 Título principal;	1A. Regras preliminares	8.1B Título principal;	Título
1.2 Designação geral do material;	1B. Título correto	8.1C Designação geral do material;	Título principal;
1.3 Subtítulo;	1C. Concepção de títulos	8.1D Títulos equivalentes;	Título paralelo;
1.4 Outras informações sobre títulos;	1D. Designação geral de material	8.1E Outras informações sobre o título;	Outras informações sobre o título;

Fonte: Simionato (2015, p. 177).

A partir dessa tabulação, os elementos que maior apresentam correspondência entre os esquemas, são: „título“, „título principal“, „título equivalente“ e „outras informações sobre o título“. (SIMIONATO, 2015, p. 112-113). Esses elementos comparados asseguram uma semântica e uma interoperabilidade ao sistema informativo.

O *crosswalk* tem um impacto direto sobre a qualidade da conversão de dados: qualquer correspondência ou falha nos pares (esquemas) pode causar conversão incorreta ou perda de milhares de valores, em outras palavras, pode gerar a inconsistência dos dados nos registros

bibliográficos, dispostos nas camadas de representação, quanto na camada de apresentação dos dados num sistema de busca e recuperação de informações.

Além do método *crosswalk*, outro fato importante a ser considerado nos metadados e esquemas de metadados, são as arquiteturas de metadados, ou seja, uma estrutura que permita comportar os metadados, de forma que tanto a estrutura, a sintaxe, quanto à semântica possam ser compreendidas por uma máquina, e ainda propiciar a interoperabilidade semântica entre ambientes e sistemas de informação heterogêneos.

Dentre as iniciativas para promover que os ambientes informativos possam comunicar-se entre si, um dos requisitos está na adoção de arquiteturas de metadados, ou seja, uma estrutura que permita integrar e alocar diferentes tipos de padrões metadados na *Web*, ou seja, garantir a interoperabilidade.

Dessa forma, as arquiteturas de metadados têm como propósito a representação, a codificação e o transporte de padrões e/ou esquemas de metadados heterogêneos com estruturas flexíveis, com o objetivo de promover a interoperabilidade nos níveis estrutural, sintático e semântico, fazendo com que os recursos informativos nos ambientes digitais possam estar cada vez mais e melhores representados e descritos.

Com a heterogeneidade de dados e metadados, disponibilizados no ambiente *Web*, faz-se necessário o desenvolvimento de um esquema ou estrutura que comporte esta diversidade de metadados, de modo que permita sistemas interoperáveis na rede. Esse esquema ou estrutura é denominado arquitetura de metadados, cujos objetivos são representar e proporcionar uma base de codificação de metadados por meio de estruturas flexíveis.

Nesse cenário, a arquitetura de metadados *ResourceDescription Framework* (RDF) utilizada de maneira interdisciplinar na Ciência da Computação e na Ciência da Informação, possibilita o processamento e o gerenciamento automatizado de recursos informativos, de acordo com áreas do conhecimento distintas, principalmente com relação: à descoberta de recursos informativos, à descrição das relações entre os recursos informativos representadas na rede, ao auxílio aos agentes de *software*, ao intercâmbio e ao compartilhamento de informações entre outros tipos de aplicações. (LASSILA; SWICK, 1999).

O RDF é um *framework* para metadados que possibilita a interoperabilidade. Ele provê uma infraestrutura que possibilita a codificação, o intercâmbio, o uso e o reuso de metadados de uma forma não ambígua, de maneira que máquinas possam compreender a semântica dos metadados e conseqüentemente usá-los no processo chamado descoberta do recurso (TAYLOR, 2004; ZENG; QIN, 2008),

promovendo um mecanismo para a integração de múltiplos esquemas de metadados para a descrição de recursos na *Web*.

O RDF também é considerado um modelo de estrutura independente de comunidades, pode desenvolver vocabulário que adaptam suas necessidades específicas e então compartilham aqueles vocabulários com outras comunidades. Descrevendo um recurso com metadados permite que ele possa ser compreendido por humanos e máquinas de forma que promova a interoperabilidade.

Ressalta-se que ao utilizar esquemas de metadados definidos, compartilhando transferência entre protocolos e o método *crosswalk* entre esquemas, os registros dos recursos bibliográficos na rede ou em catálogos podem ser buscados mais amplamente. De tal modo que a essência de um sistema de modelagem de dados seja, identificar o que é necessário para um modelo de conceituar algum domínio é a maneira de denotar entidades e referenciá-las de forma inequívoca.

Nesse sentido, cada ontologia deve fornecer os significados para designar entidades. Existindo entidades que podem ser referenciadas, o próximo passo é descrever as relações entre elas. Como as relações são elementos fundamentais semânticos, devem também ser inequivocamente endereçável. As propriedades em RDF podem ser vistas como relações binárias. Este é o tipo mais básico de relações entre duas entidades. Os tipos de relações mais complexas podem ser modelados através da definição de um vocabulário para as propostas no topo de RDF, tal como tem sido feito em *Ontology Web Language* (OWL). (HEATH; BIZER, 2011).

As várias linguagens de ontologia diferem no seu vocabulário, seus fundamentos lógicos e nos seus elementos epistemológicos, mas elas têm em comum o fato de descrever as estruturas das entidades e suas relações. Portanto, RDF é o maior denominador comum de todas as linguagens de ontologia. (HEATH; BIZER, 2011). RDF não é apenas uma forma de codificar as linguagens de ontologia ou apenas um modelo de dados arbitrário, mas é um modelo de dados estruturado que corresponde exatamente à estrutura das linguagens de ontologia e à codificação de elementos de metadados.

No domínio bibliográfico, as comunidades têm investigado orientações que podem ser potencialmente utilizadas em harmonia com o RDF, tal como o *Resource Description & Access* (RDA), projetado para a representação e a descrição de recursos informativos em meio digital e compatível com a *Web Semântica*.

Resource Description & Access (RDA) é um novo padrão para descrever metadados de recursos mantidos nas coleções de bibliotecas, arquivos, museus e outras organizações de gerenciamento de informações. Desde 2007, o grupo de trabalho do padrão

ResourceDescription and Access (RDA) tem realizado um esforço paralelo para construir vocabulários habilitados no contexto da *Web Semântica*.

Seu objetivo é fornecer um conjunto compreensível de diretrizes e instruções descritivas para a criação de metadados oriundos de todos os tipos de mídias e conteúdo de recursos. Suas orientações estão baseadas no código de catalogação AACR2r e no desenvolvimento dos Requisitos Funcionais para Registros Bibliográficos (FRBR) e também na Declaração Internacional de Princípios de Catalogação, formulada na primeira década do século XXI.

Os ambientes informacionais digitais ao adotarem a arquitetura de metadados RDF, juntamente com ontologia e os esquemas de metadados podem encontrar um caminho para o estabelecimento da interoperabilidade semântica entre sistemas heterogêneos e distintos.

Os agentes propostos para a interoperabilidade semântica são os denominados „metadados semânticos“ (*semantic metadata*), que se referem a anotações legíveis por máquinas (*machine-processable annotations*) utilizando a arquitetura RDF, que no contexto da *Web Semântica*, entende-se como o processo de atribuição de termos de uma ontologia a um recurso, independentemente de sua tipologia, ou de uma parte do recurso, por exemplo, uma palavra ou uma sentença. (BÜRGER; SIMPERL; TEMPICH, 2014).

O uso dos modelos conceituais de dados, das arquiteturas de metadados e das ontologias redesenham os novos ambientes informacionais digitais, definindo conceitualmente os elementos da descrição bibliográfica a serem representados pelo catalogador, proporcionando interfaces de buscas mais compreensíveis aos usuários e no estabelecimento efetivo da interoperabilidade.

3 Considerações finais

A ontologia e os metadados encontram-se como bases estruturantes das descrições bibliográficas atuais e podem ser consideradas como elementos básicos e norteadores para o estabelecimento da interoperabilidade semântica entre os ambientes e sistemas informacionais heterogêneos.

Nesse sentido, pode-se corroborar com a assertiva de Jacob (2003) ao se referir que todos os esquemas de metadados estão atrelados à ontologia, onde procuram especificar um conjunto de elementos (características físicas e/ou conceituais) de um recurso, julgadas relevantes para uma determinada comunidade ou área de conhecimento particular.

Considera-se necessária uma rearticulação dos conceitos e bases epistemológicas que abarcam a ontologia e os metadados, na tentativa de inseri-los em domínios específicos de aplicação, como a Ciência da

Informação, ou sua utilização para fins mais abrangentes, potencializando suas funcionalidades e permeadas pela dinâmica da evolução das TIC.

Retoma-se o pensamento de que a heterogeneidade de recursos disponibilizados no ambiente *Web* atualmente requer o uso estratégico e efetivo do instrumental que pode contribuir para o estabelecimento e para a modelagem dos catálogos bibliográficos de forma mais contundente, tais como a ontologia e os metadados, proporcionando aos sistemas de informação a ideia elementar de melhores formas de encontrar, identificar, selecionar e obter recursos informativos e/ou registros bibliográficos no atendimento às necessidades dos usuários.

A ontologia e os metadados utilizados de forma estratégica e sinérgica poderão garantir a integridade das informações, fornecendo elementos de descrição mais ricos semanticamente e proporcionando ainda os relacionamentos entre um conjunto de esquemas, oriundos de ambientes informativos heterogêneos, garantindo à interoperabilidade semântica e multidimensionando as formas de acesso aos recursos, para posterior utilização e reutilização de informações.

Dessa forma, é possível observar contemporaneamente o trabalho conjunto interdisciplinar almejando a interoperabilidade semântica, no investimento nas camadas estruturantes e determinantes dos ambientes informativos digitais, entendidas pela representação e descrição dos dados, principalmente com a utilização de esquemas de metadados e ontologia que definem os relacionamentos, potencializam a busca e o acesso e garantem a recuperação da informação de forma mais efetiva.

Referências

ALMEIDA, M. B. Revisiting ontologies: a necessary clarification. *Journal of the American Society for Information Science and Technology*, [s.l.], v. 64, n. 8, p. 1682-1693, 2013.

ALMEIDA, M. B. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. *Perspectivas em Ciência da Informação*, Belo Horizonte, v.19, n.3, p.242-258, jul./set. 2014.

ALVES, R. C. V. *Web Semântica: uma análise focada no uso de metadados*. 2005. 180 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia Ciências, Universidade Estadual Paulista - UNESP, Marília, 2005.

BACA, M. *Introduction to metadata 3.0*. 2. ed. Los Angeles: The Getty Research Institute, 2008. Disponível em: http://www.getty.edu/research/conducting_research/standards/intrometadata/pdf.html. Acesso em: 20 maio 2017.

BÜRGER, T.; SIMPERL, E.; TEMPICH, C. Methodologies for the creation of semantic data. In: SICILIA, M. A. *Handbook of metadata, semantics and ontologies*. London: World Scientific Publishing Co, 2014. p. 185-216.

CASTRO, F. F. de; SANTOS, P. L. V. A. C. Os metadados como instrumentos tecnológicos na padronização e potencialização dos recursos informativos no âmbito das bibliotecas digitais na era da web semântica. *Informação & Sociedade: Estudos*, João Pessoa, v. 17, n. 2, p. 13-19, maio/ago. 2007.

CASTRO, F. F. de. *Padrões de representação e descrição de recursos informativos em bibliotecas digitais na perspectiva da ciência da informação: uma abordagem do MarcOnt initiative na era da web semântica*. 2008. 201 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista - UNESP, Marília, 2008.

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

CHATEAUBRIAND, O. A filosofia, a linguagem e o mundo. In: BRITO, A. N. de; VALE, O. A. *Filosofia, linguística, informática: aspectos da linguagem*. Goiânia: Ed. UFG, 1998.

CORCHO, O.; FERNÁNDEZ-LOPEZ, M.; GÓMEZ-PÉREZ, A. OntoWeb: ontology-based information exchange for knowledge management and electronic commerce. *Technical RoadMap*, v. 1.0. Madrid: Universidad Politécnica de Madrid, 2001. Disponível em: <https://docplayer.net/2209059-Ontoweb-ontology-based-information-exchange-for-knowledge-management-and-electronic-commerce-ontoweb-workshop.html>. Acesso em: 23 jan. 2017.

CRUZ, I. F.; XIÃO, H. The role of ontologies in data integration. *Journal of Engineering Intelligent Systems*, [s.l.], v. 13, n. 4, 2005. Disponível em: <http://www2.cs.uic.edu/~advis/publications/dataint/eis05j.pdf>. Acesso em: 10 abr. 2017.

CUNHA, L. M. S. *Web semântica: estudo preliminar*. Campinas: Embrapa Informática Agropecuária, 2002. Disponível em: <http://www.cnptia.embrapa.br/modules/tinycontent3/content/2002/doc18.pdf>. Acesso em 03 fev. 2017.

GREENBERG, J.; SUTTON, S.; CAMPBELL, D. Metadata: a fundamental component of the semantic web. *Bulletin of the American Society for Information Science*, [s.l.], v. 29, n. 4, p.16-18, 2003.

GRUBER, T. What is an ontology? *Knowledge Acquisition*, [California], v. 5, n. 2, p. 199-220, 1993. Disponível em: <http://http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>. Acesso em: 10 jul. 2017.

HEATH, T.; BIZER, C. *Linked data: evolving the web into a global data space*. [S.l.]: Morgan & Claypool, v. 1 n. 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 jun. 2017.

JACOB, E. K. Ontologies and the semantic web. *Bulletin of the American Society for Information Science and Technology*, [s.l.], v. 29, n. 4, p. 19-22, Apr./May, 2003. Disponível em: <http://www.asis.org/Bulletin/Apr-03/jacob.html>. Acesso em: 27 abr. 2017.

LASSILA, O.; SWICK, R. R. *Resource description framework (RDF) model and syntax specification*. [S. l.: s. n.], 1999. Disponível em: <http://www.w3.org/TR/1999/REC-rdfsyntax-19990222/>. Acesso em: 3 jul. 2017.

LUBAS, R.; JACKSON, A.; SCHNEIDER, I. *The metadata manual: a practical workbook*. Sawston: Elsevier, 2013.

MILSTEAD, J.; FELDMAN, S. *Metadata: cataloging by any other name*. [S. l.: s. n.], 1999.

RILEY, Jenn. *Understanding metadata*. Bethesda, USA: National Information Standards Organization, 2004. 20 p. Disponível em: <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>. Acesso em: 10 abr. 2017.

OWL. *Web Ontology Language: use cases and requirements*. Pennsylvania: W3C, 2004. Disponível em: <http://www.w3.org/TR/webont-req/#onto-def>. Acesso em: 27 abr. 2017.

SARACEVIC, T. Interdisciplinary nature of information science. *Ciência da Informação*, Brasília, DF, v. 24, n. 1, p. 36-41, 1995.

SENSO, J. A.; ROSA PIÑERO, A. El concepto de metadato: algo más que descripción de recursos electrónicos. *Ciência da Informação*, Brasília, DF, v. 32, n. 2, p. 95-106, 2003. Disponível em: <http://www.scielo.br/pdf/ci/v32n2/17038.pdf>. Acesso em: 10 jul. 2017.

SIMIONATO, A. C. *Modelagem conceitual DILAM: princípios descritivos de arquivos, bibliotecas e museus para o recurso imagético digital*. 2015. 200 f. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Marília, 2015.

SIQUEIRA, M. A. *XML na ciência da informação: uma análise do MARC21*. Marília, 2003, 134 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista - UNESP, Marília, 2003.

SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, Brasília, DF, v. 33, n. 1, p. 132-141, jan./abr. 2004.

TAYLOR, A. G. *The organization of information*. 2. nd. Westport: Libraries Unlimited, 2004. 417 p. (Library and Information Science Text Series).

VELLUCCI, S. L. Metadata and authority control. *Library Resources and Technical Services*. [S.l.], v. 44, n. 1, p. 33-43, Jan. 2000.

VICKERY, B. C. Ontologies. *Journal of Information Science*, London, v. 23, n. 4, p. 227-286, 1997.

ZENG, M. L.; QIN, J. *Metadata*. New York: Neal-Schuman Publishers, 2008. 365 p.