

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

PAULO GEORGE MIRANDA MARTINS

EVOLUÇÃO DAS TECNOLOGIAS DE REPRESENTAÇÃO:
das linguagens de marcação aos dados interligados

São Carlos
2018

PAULO GEORGE MIRANDA MARTINS

EVOLUÇÃO DAS TECNOLOGIAS DE REPRESENTAÇÃO: das linguagens de
marcação aos dados interligados

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos - Campus de São Carlos, para obtenção do título de Mestre em Ciência da Informação.

Área de Concentração: Conhecimento, Tecnologia e Inovação.

Linha de pesquisa: Tecnologia, Informação e Representação.

Orientador: Prof. Dr. Rogério Aparecido Sá Ramalho

São Carlos - SP
2018



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência da Informação

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Paulo George Miranda Martins, realizada em 08/08/2018:



Prof. Dr. Rogério Aparecido Sá Ramalho
UFSCar

Prof. Dr. Ricardo César Gonçalves Sant'Ana
UNESP



Profa. Dra. Ana Carolina Simionato
UFSCar

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Ricardo César Gonçalves Sant'Ana e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ão) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.



Prof. Dr. Rogério Aparecido Sá Ramalho

DEDICATÓRIA

Dedico à minha amada Sandra pelo apoio, dedicação e companheirismo de todos os momentos e, ao nosso pequeno Miguel com todo meu amor.

AGRADECIMENTO

A Deus por TUDO que tem me proporcionado, dando-me saúde e força para superar todos os obstáculos impostos pela vida.

Agradeço aos meus amados pais, Antônio Valdir e Maria Inês por serem à base de tudo para mim e pelos ensinamentos que me orientaram para o caminho do bem.

À minha amada esposa Sandra, a grande responsável pela realização desse sonho pela dedicação, carinho, incentivo, companheirismo e acima de tudo por ser minha inspiração de humildade e vida. Sem você Vida nada disso seria possível.

Ao meu orientador Rogério Ramalho, pela paciência, dedicação, amizade e profissionalismo. Aos professores Ana Carolina Simionato e Ricardo Sant'Ana, pelo profissionalismo e contribuições para o enriquecimento da pesquisa. Ao pessoal do departamento, Mônica, Renan e o Artur pela amizade e prestatividade com as nossas solicitações.

A todos os professores do Programa de Pós-Graduação em Ciência da Informação (PPGCI), pelos exemplos de profissionalismo e dedicação para o que o programa se tornasse uma realidade.

Aos meus amigos Helton Graciano e Fábio Renato Rodrigues pelo incentivo e, aos meus queridos amigos de mestrado pela amizade para a vida toda: Laila, Débora, Raquel, Tamie, Jéssica, Melina, Humberto, Júlio e Guilherme.

Enfim, agradeço a todos aqueles que direta ou indiretamente contribuíram para a realização desta pesquisa. O meu MUITO OBRIGADO.

RESUMO

As denominadas Linguagens de Representação sempre tiveram relação direta com os instrumentos e práticas profissionais identificados no campo da Ciência da Informação. A partir do desenvolvimento das tecnologias digitais os formatos bibliográficos de representação têm evoluído buscando atender às demandas informacionais emergentes. Nesse sentido, os métodos de representação sempre foram objetos de investigação no campo da Ciência da Informação que desde a sua criação, desenvolve métodos voltados para a representação dos diversos tipos de recursos informacionais. Diante deste cenário, questiona-se: quais as relações existentes entre as linguagens de representação convencionalmente utilizados na área de Ciência da Informação e as tecnologias de publicação de dados interligados? O objetivo da pesquisa é analisar a evolução das linguagens de representação e identificar as relações existentes entre os instrumentos de representação utilizados na área de Ciência da Informação e as novas tendências de publicação de dados interligados. Trata-se de uma pesquisa teórica de abordagem qualitativa, e natureza aplicada, focada na observação analítica dos documentos que abordam a temática de representação da informação. Utiliza-se como procedimentos metodológicos a análise bibliográfica e documental, abordando o volume de dados e informações obtidos mediante a técnica de análise de conteúdo. A pesquisa justifica-se devido à necessidade de estudos que corroborem para um melhor entendimento dos novos modelos de representação fundamentados na proposta de dados interligados e suas aplicações para a área de Ciência da Informação. Como resultado, é apresentada uma visão geral da evolução das linguagens de representação e descritas as principais relações com as novas tendências de publicação de dados interligados, com destaque para o modelo de dados BIBFRAME.

Palavras-chave: Linguagens de Marcação. Modelo de Dados. Representação da Informação. Tecnologias Semânticas. BIBFRAME.

ABSTRACT

The so-called Representation Languages have always been directly related to the instruments and professional practices identified in the field of Information Science. From the development of digital technologies, the bibliographic representation formats have evolved to meet emerging information demands. In this sense, the methods of representation have always been objects of investigation in the field of Information Science that since its creation, develops methods directed to the representation of the diverse types of informational resources. In view of this scenario, we ask: what are the relations between the representation languages conventionally used in the area of Information Science and the technologies of interconnected data publishing? The objective of the research is to analyze the evolution of representation languages and to identify the existing relationships between the representation tools used in the area of Information Science and the new trends of interconnected data publication. It is a theoretical research of qualitative approach, and applied nature, focused on the analytical observation of the documents that approach the thematic of information representation. Bibliographic and documentary analysis is used as methodological procedures, addressing the volume of data and information obtained through the technique of content analysis. The research is justified due to the need for studies that corroborate for a better understanding of the new models of representation based on the proposal of interconnected data and its applications for the area of Information Science. As a result, an overview of the evolution of representation languages is presented and the main relationships with the new interrelated data publishing trends are described, especially the BIBFRAME data model.

Keywords: Markup Languages. Data Model. Representation of Information. Semantic Technologies. BIBFRAME.

LISTA DE FIGURAS

Figura 1 - Evolução da Web em uma Web de dados interligados.....	24
Figura 2 - Exemplo de documento HTML.....	30
Figura 3 - Exemplo básico de um catálogo em XML.....	37
Figura 4 - Fragmento de XML Schema.....	38
Figura 5 - Cronologia de desenvolvimento das LMs.....	40
Figura 6 - Exemplo básico de um Grafo da declaração RDF.....	44
Figura 7 - Grafo RDF de triplas simples.....	45
Figura 8 - Estrutura de RDF/XML.....	47
Figura 9 - Relações entre declarações RDF.....	48
Figura 10 - Declaração RDF.....	49
Figura 11 - Relações sintáticas da OWL.....	52
Figura 12 - Estrutura da OWL 2.....	53
Figura 13 - Esquema MARCXML simplificado.....	64
Figura 14 - Representação do Dublin Core em RDF.....	70
Figura 15 - Amostra de metadados do DC para uma página da web.....	72
Figura 16 - Modelo BIBIFRAME 1.0.....	76
Figura 17 - Modelo BIBIFRAME 2.0.....	77
Figura 18 - Principais Classes do BIBIFRAME e do Grupo 1 do FRBR.....	79
Figura 19 - Modelo de dados BIBIFRAME com mapeamento MARC e RDA.....	80

LISTA DE QUADROS

Quadro 1 - Versões da linguagem HTML.....	31
Quadro 2 - Especificações das sintaxes da OWL	54
Quadro 3 - Especificações da OWL 2	55
Quadro 4 - Atributos das Linguagens de Representação.....	56

LISTA DE ABREVIATURAS E SIGLAS

ANSI	<i>American National Standards Institute</i>
API	<i>Application Programming Interface</i>
BDTD	<i>Biblioteca Digital de Teses e Dissertações</i>
BIBFRAME	<i>Bibliographic Framework</i>
CERN	<i>European Organization for Nuclear Research</i>
CI	<i>Ciência da Informação</i>
DC	<i>Dublin Core</i>
DCMI	<i>Dublin Core Metadata Initiatives</i>
DCMI-TERMS	<i>Metadata Terms</i>
DCAM	<i>DCMI Abstract Model</i>
DTD	<i>Document Type Definition</i>
GML	<i>Generalized Markup Language</i>
HTML	<i>HyperText Markup Language</i>
IRI	<i>International Resource Identifier</i>
ISKO	<i>International Society for Knowledge Organization</i>
LC	<i>Library of Congress</i>
LD	<i>Linguagens Documentárias</i>
LM	<i>Linguagens de Marcação</i>
LN	<i>Linguagem Natural</i>
METS	<i>Metadata Encoding and Transmission Standard</i>
MODS	<i>Metadata Object Description Schema</i>
NDLTD	<i>Networked Digital Library of Theses and Dissertations</i>
OAI	<i>Open Archive Initiative</i>
OAI-PMH	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>
OC	<i>Organização do Conhecimento</i>
OPAC	<i>Online Public Access Catalog</i>
ORC	<i>Organização e Representação do Conhecimento</i>
ORI	<i>Organização e Representação da Informação</i>
OWL	<i>Ontology Web Language</i>
RDF	<i>Resource Description Framework</i>
SciELO	<i>Scientific Electronic Library Online</i>

SGML	<i>Standard Generalized Markup Language</i>
TIC	Tecnologias da Informação e Comunicação
UNISIST	<i>United Nations International Scientific Information System</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
W3C	<i>World Wide Web Consortium</i>
XHTML	<i>eXtensible HyperText Markup Language</i>
XML	<i>eXtensible Markup Language</i>
XSD	<i>XML Schema Definition</i>

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Problema de pesquisa.....	15
1.2 Proposição	16
1.3 Justificativa.....	16
1.4 Objetivos	17
1.5 Procedimentos metodológicos	18
1.6 Estrutura da dissertação	18
2 DOS CATÁLOGOS ONLINE AOS DADOS INTERLIGADOS	20
3 LINGUAGENS DE REPRESENTAÇÃO	26
3.1 <i>Standard Generalized Markup Language</i> (SGML)	27
3.2 <i>Hypertext Markup Language</i> (HTML).....	29
3.3 <i>eXtensible Markup Language</i> (XML).....	35
3.4 <i>Resource Description Framework</i> (RDF)	41
3.5 <i>Web Ontology Language</i> (OWL)	50
4 EVOLUÇÃO DOS FORMATOS BIBLIOGRÁFICOS DE REPRESENTAÇÃO	58
4.1 Formato MARC	60
4.2 <i>Dublin Core</i>	67
4.3 <i>Bibliographic Framework</i> (BIBFRAME)	72
5 CONSIDERAÇÕES FINAIS	84
5.1 Pesquisas Futuras	87
REFERENCIAS	89

1 INTRODUÇÃO

As denominadas Linguagens de Representação sempre tiveram relação direta com os instrumentos e práticas profissionais identificados no campo da Ciência da Informação. A partir do desenvolvimento do ambiente *Web* tais tecnologias ganharam maior destaque, impulsionando a criação de uma grande variedade de aplicações com diferentes características e propósitos.

Ao longo das últimas décadas o grande volume de informações produzidas, associado ao desenvolvimento das Tecnologias de Informação e Comunicação (TIC), tem cada vez mais potencializado estudos direcionados para a criação e desenvolvimento de novos modelos e métodos de representação informacional no âmbito da área de Ciência da Informação, que visa permitir uma maior capacidade de armazenamento de recursos com vistas a possibilitar uma recuperação eficiente, analisando os impactos e consequências que permeiam todas as etapas e processos inerentes aos fluxos informacionais.

Em um contexto social com grandes transformações e evoluções tecnológicas, o uso de modelos clássicos no processo de representação da informação necessita ser reformulado sob uma nova perspectiva, pois,

[...] para a garantia de qualidade na recuperação, apresenta-se como desafio a necessidade de singularização contextual na reconstrução do conhecimento, a partir da determinação de requisitos de qualidade e relevância das informações, que permitam categorizar e organizar, de maneira eficiente, o “oceano” de dados disponíveis, favorecendo a identificação de informações que realmente interessam ao usuário (RAMALHO; VIDOTTI; FUJITA, 2007, p. 1).

A informação sempre foi um insumo importante para o desenvolvimento humano e o processo de representação permite a construção de novos conhecimentos através do compartilhamento do conteúdo representado, na qual a criação de “[...] um sistema de recuperação de informação tem como ponto focal o usuário e suas necessidades de informação” (BOCCATO, 2011, p. 23).

Assim,

As formas de publicação, disseminação, comunicação e recuperação da informação se alteraram com as diversas possibilidades ofertadas pelo mundo virtual. Assim, representar a informação tornou-se uma atividade mais complexa, tendo em vista que os sistemas automatizados conseguem realizar combinações e conexões simultâneas no processo de busca (FERREIRA, 2013, p. 2).

À medida que novas tecnologias têm sido apresentadas, o compartilhamento de conhecimento entre diversos grupos sociais tem aumentado paulatinamente nas últimas décadas e, com o avanço no desenvolvimento dos primeiros computadores eletrônicos sua aplicação no processamento de informações bibliográficas possibilitou uma maior capacidade de representação, armazenamento e organização de informações, cuja finalidade é a recuperação independente dos aportes tecnológicos em que esteja armazenada (MEADOWS, 1998; CAPURRO, 2003; LE COADIC, 2004).

Desde seu surgimento, a área de Ciência da Informação tem por objetivo principal “[...] reunir, organizar e tornar acessível o conhecimento cultural, científico e tecnológico produzido em todo o mundo” (OLIVEIRA, 2008, p. 13). A CI é a área do conhecimento que se ocupa em estudar e propor métodos e processos de organização e representação de informações, possibilitando que as necessidades informacionais dos indivíduos sejam atendidas em seu contexto social, cultural e/ou político (BARRETO, 2002; ORTEGA, 2013).

Nesse contexto, o ambiente *Web* é uma das mais ricas fontes de informação (RAMALHO, 2015) e diversos aplicativos criados para esse ambiente fornecem meios de acesso rápido aos dados disponíveis que vão desde uma simples visualização até ferramentas que permitam a integração e processamento de dados na escala da *Web*.

No entanto, apesar de todas as tecnologias desenvolvidas a diversidade de padrões e modelos usados para a representação pode resultar em processos ineficientes caso não possuam uma estrutura semântica compatível com os novos mecanismos que visam tornar o intercâmbio de informações, no meio digital, um processo mais eficiente.

A criação de modelos de representação para ambientes digitais traduz-se em um esforço no sentido de converter a *Web* em uma *Web Semântica* e a representação de informações em um sistema digital devem ser obtidas a partir de uma estrutura navegável que permita o relacionamento de entidades e atributos entre os diversos recursos informacionais disponíveis (BISCHOF et al., 2011).

Para Catarino (2014, p. 19), a “[...] *Web Semântica* não é uma “nova *Web*” e sim uma proposta de mudanças para a *Web*, na qual a informação possui um significado bem definido, permitindo melhor interação entre computadores e

peças”. Ou seja, a partir da criação de tecnologias voltadas para as inferências de novos recursos com os já disponíveis, tem-se uma maior capacidade de se interligar partes de conteúdos através do uso de estruturas com maior capacidade semântica no qual o usuário de sistemas informatizados faz parte do processo de enriquecimento de conceitos, permitindo que esses sistemas proporcionem maior amplitude na recuperação de informações.

Dessa maneira,

[...] as formas de organizar e de recuperar sempre estiveram relacionadas às tecnologias vigentes, de modo que a constante evolução das tecnologias e o avanço exponencial na quantidade de informação disponível, principalmente no meio digital, vêm conduzindo a um estágio em que os modelos clássicos de organização e recuperação de informações precisam ser (re) pensados sob diferentes perspectivas (RAMALHO, 2006, p. 37).

A partir do desenvolvimento das tecnologias digitais os formatos bibliográficos de representação têm evoluído buscando atender às demandas informacionais emergentes. No entanto, o principal desafio para o desenvolvimento de modelos de representação é a criação de padrões com estruturas que possibilitem agregar maior semântica dos dados representados quanto a sua terminologia, pois a ausência desta:

[...] tende a prejudicar a consecução da interoperabilidade semântica entre os repositórios e também com outros sistemas de informação, em razão das inconsistências semânticas – diferentes terminologias e significado – e dos esquemas de metadados (ANDRADE; CERVANTES, 2012, p. 154).

Entretanto, ainda há limitações quanto ao tratamento de todo o conteúdo informacional disponível na *Web* (VICTORINO; MEDEIROS; SANTOS, 2011), uma vez que a estrutura de formatos de representação tradicionais, como o MARC, nem sempre suporta as mesmas especificidades de entradas de modelos que utilizam uma semântica mais ampla, o que, em determinados casos, não possibilita um compartilhamento de informação entre sistemas.

A existência de um padrão compatível entre sistemas operacionais “[...] previne a duplicidade de trabalho e possibilita o compartilhamento de registros bibliográficos” (ZAFALON, 2008, p. 15). Como resultado da padronização dos processos de representação, tem-se a facilidade e qualidade no acesso e, a eficácia na recuperação de conteúdos feita por diversos tipos de usuários (ANDRADE;

CERVANTES, 2012).

Com a popularização da *Web*, no final da década de 1990, a mesma tem-se transfigurado de uma rede de *hiperlinks* para uma rede de dados e documentos interligados. Nesse cenário, destaca-se o desenvolvimento de modelos que potencializam o intercâmbio de informações a partir da interligação de dados disponíveis em domínios distintos, como o *Resource Description Framework* (RDF), uma estrutura simples que utiliza a linguagem XML para descrever recursos informacionais a partir dos relacionamentos entre outros recursos já existentes e, o BIBFRAME, uma iniciativa da LC que favorece o desenvolvimento de padrões de descrição bibliográfica e possibilita o relacionamento entre diversos recursos através de *links* na *web* (RAMALHO, 2016).

Nessa perspectiva, dada a ampla discussão na área de CI sobre as tecnologias de representação desenvolvidas para ambientes digitais e suas aplicações, observa-se a necessidade de estudos que favoreçam subsídios teóricos para um melhor entendimento sobre as tecnologias semânticas utilizadas na representação de informações, seus conceitos e métodos de análise e, síntese possibilitando que tais avanços possam ser devidamente compreendidos pelos profissionais da informação e sedimentados no campo teórico da área, identificando as principais características e impactos destas tecnologias na área de Ciência da Informação.

1.1 Problema de pesquisa

Diante do imenso volume informacional disponível em ambientes digitais e considerando as novas ferramentas tecnológicas, o uso de modelos de representação da informação alinhados com as propostas emergentes no contexto de dados interligados evita o retrabalho e permite o reuso de conteúdos já desenvolvidos para ambientes digitais, permitindo maior compartilhamento e recuperação de informações.

Nessa perspectiva, com enfoque específico na temática evolução das linguagens e tecnologias de representação bibliográfica, surge a seguinte questão de pesquisa: quais as relações existentes entre as linguagens de representação convencionalmente utilizadas na Ciência da informação e as tecnologias de

publicação de dados interligados?

1.2 Proposição

A presente pesquisa parte da necessidade de compreender o uso das linguagens de representação e sua aplicação no desenvolvimento de modelos e formatos de representação bibliográfica para o contexto *Web*. Nesse sentido, propõe-se analisar as questões que norteiam a evolução de tais tecnologias e suas aplicações em ambientes digitais.

Para tanto, é requerido um estudo aprofundado dos princípios e conceitos das linguagens e formatos de representação bibliográfica e, das novas tecnologias semânticas alinhadas com os propósitos de publicação de dados vinculados.

1.3 Justificativa

A partir da proposição deste trabalho, a justificativa delinea-se pelo aprofundamento nos estudos sobre os modelos de representação iniciado no Trabalho de Conclusão de Curso, cuja abordagem dava-se na análise das linguagens de marcação e padrões de metadados utilizados para a representação de informações, tendo como objeto de estudo o *Metadata Authority Description Schema* (MADS).

Em um universo em que a conectividade e o intercâmbio de informações estão cada vez mais presentes em nosso cotidiano, a diversidade de modelos de representação disponíveis pode constituir um problema dado o elevado número de documentos produzidos para ambientes digitais e a incoerência terminológica de determinados recursos disponíveis na *Web*.

Nesse sentido, conhecer as potencialidades oferecidas pelas linguagens de representação é fundamental para se compreender o desenvolvimento de novas tecnologias semânticas fundamentadas na proposta de dados interligados, no qual o ponto focal desse movimento é possibilitar um maior relacionamento entre recursos informacionais, favorecendo o compartilhamento e a recuperação de informações de modo eficiente, atendendo as necessidades informacionais dos usuários de forma ágil e precisa.

No campo científico, espera-se que o estudo fomente discussões na área da Ciência da Informação sobre o desenvolvimento das temáticas em questão, possibilitando a indicação de criação de novos modelos de representação informacional uma vez que, observou-se um número limitado de pesquisas voltadas para a temática que envolve formatos de representação para publicação de dados interligados.

Na área social, espera-se que o estudo proporcione subsídios teóricos para os profissionais da informação e biblioteconomia sobre as novas tecnologias e suas aplicações no âmbito da Ciência da Informação. Destaca-se que, com um conhecimento mínimo sobre essas tecnologias os profissionais da informação possam propor a adoção de modelos e/ou formatos que melhor atendam suas necessidades levando em consideração seu público, instituições parceiras e compartilhamento de conteúdo informacional.

1.4 Objetivos

O objetivo da pesquisa é analisar a evolução das linguagens de representação e identificar as relações existentes entre os instrumentos de representação utilizados na área e Ciência da Informação e as novas tendências de publicação de dados interligados.

Os objetivos específicos são:

- apresentar os conceitos e métodos acerca da criação das linguagens de representação a partir da análise da conjuntura científica e literária ao longo das últimas décadas;
- destacar a relação entre as práticas tradicionais e as novas tecnologias de representação;
- destacar a aplicação e contribuição das linguagens de representação para a área de Ciência da Informação;
- destacar os aspectos conceituais dos formatos de representação bibliográfica (sintaxe e semântica) e suas aplicações a partir da proposta de dados interligados na *web*.

1.5 Procedimentos metodológicos

Caracteriza-se como uma pesquisa de abordagem qualitativa e de natureza aplicada, uma vez que busca avaliar e relacionar a questão de pesquisa aos objetivos específicos propostos. Quanto ao método adotado, a pesquisa qualitativa “[...] busca compreender e analisar a realidade dos enfoques da pesquisa” (TREVIÑOS, 1987, p. 117).

No que diz respeito aos objetivos propostos para esse estudo, a pesquisa é exploratória e descritiva buscando o conhecimento, a compreensão e a “[...] familiaridade com o objeto a ser investigado” (GONDIM; LIMA, 2006, p. 70).

Para a construção de um arcabouço teórico robusto, fez-se um levantamento de referencial teórico, bibliográfico e documental, nas bases de dados relacionados ao tema estudado, publicações nacionais e internacionais nos idiomas, português, espanhol e inglês, na aquisição de livros, periódicos, artigos científicos, dissertações, teses e consulta a sites que abordam temas relacionados às linguagens de representação, formatos de representação bibliográfica e as tecnologias semânticas fundamentadas na proposta de dados interligados.

A localização dos documentos necessários (artigos, trabalhos apresentados em eventos, relatórios científicos, dissertações, teses e livros), deu-se através de consultas nas seguintes bases de dados: Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (Brapci), Portal de Periódicos da Capes, *Scientific Electronic Library Online* (SciELO), *Networked Digital Library of Theses and Dissertations* (NDLTD), Biblioteca Digital de Teses e Dissertações (BDTD), *World Wide Web Consortium* (W3C) e na *Library of Congress* (LC).

1.6 Estrutura da dissertação

A estrutura para a elaboração dessa pesquisa apresenta no capítulo 1, além da introdução que aborda as questões iniciais e a contextualização do tema proposto, a definição da questão de pesquisa, a motivação, a justificativa, o objetivo geral e os específicos e, os procedimentos metodológicos adotados.

O capítulo 2 aborda o aspecto transitório dos catálogos online para a era dos dados interligados. No capítulo 3, aborda-se a temática acerca das Linguagens de Representação e suas aplicações no âmbito da Ciência da Informação, destacando

a criação da primeira linguagem, a *Generalized Markup Language* (GML), até sua especificação para o padrão ISO 8879/86 ou *Standard Generalized Markup Language* (SGML). Destaca-se ainda a importância, para a WWW, da criação do *HyperText Markup Language* (HTML) e, posteriormente baseada em SGML, a metalinguagem *eXtensible Markup Language* (XML) que é mundialmente usada para a criação de outras linguagens além de ser adotada na estruturação de novos modelos e padrões de representação de metadados, como é o caso do MARCXML e do *Dublin Core*.

Já o *Resource Description framework* (RDF) representa um novo conceito na forma de representação da informação na *Web* a partir do intercâmbio eficaz de dados entre processos automatizados e, por fim destaca-se a *Web Ontology Language* (OWL) uma linguagem que permite a criação de ontologias semanticamente complexas e ricas para se trabalhar na *Web*.

No capítulo 4, apresenta-se a evolução dos formatos bibliográficos de representação destacando o formato MARC e suas atualizações ao longo dos anos até o MARCXML. Destaca-se também o *Dublin Core*, um formato conciso, no que diz respeito ao número de elementos em sua estrutura, desenvolvido para o dinamismo exigido pela *Web* e, finaliza-se esse tópico com o *Bibliographic Framework* ou BIBFRAME um modelo fundamentado na proposta de dados interligados.

Por fim, apresenta-se no capítulo 5 as considerações finais da pesquisa seguido da proposta de pesquisas futuras e das referências utilizadas para o constructo teórico deste trabalho.

2 DOS CATÁLOGOS ONLINE AOS DADOS INTERLIGADOS

As bibliotecas têm uma longa e rica história de uso de tecnologias de representação, na qual entre as tecnologias inovadoras utilizadas em bibliotecas pode-se destacar o surgimento dos Catálogos On-line de Acesso Público (*Online Public Access Catalogs – OPAC*). Os catálogos, considerados um dos principais meios de interação do usuário para recuperação das informações armazenadas em bibliotecas, com o desenvolvimento tecnológico passaram a ser disponibilizados na *Web*.

Com a popularização da *Web* (BERNERS-LEE, 1998), no final da década de 1990, os Catálogos On-line passaram a ser desenvolvidos utilizando linguagens como HTML e ofereceram uma alternativa viável para a disseminação dos acervos e conteúdos disponíveis nas bibliotecas.

Contudo, duas décadas após a popularização da *Web*, novas tecnologias foram desenvolvidas e atualmente a rede global de computadores não se limita a disponibilização de documentos, possibilitando representar praticamente qualquer entidade do mundo real que possa ser associada a um identificador, contribuindo para inúmeras transformações no campo da representação ao adotar tecnologias com grande capacidade de compartilhamento de informações na Internet.

Tais transformações convergem para um novo cenário, no qual a comunicação e o conhecimento ocorrem em uma velocidade sem precedente na história. Contudo, para que todo o conhecimento gerado ao longo das décadas seja disseminado sistemicamente, a representação deve ser realizada com o auxílio de sistemas estruturados que permitam uma recuperação informacional eficiente, independente do sistema no qual a informação esteja armazenada.

Assim, o processo de representação tem como um de seus principais objetivos, tornar uma linguagem complexa em um documento simples, compreensível e acessível (NOVELLINO, 1996), dando ao usuário final rapidez na recuperação do conteúdo pesquisado.

Nesse cenário, surge “[...] uma grande variedade de conceitos e construtos teóricos, empíricos e pragmáticos, bem como numerosas realizações práticas” para esse processo (SARACEVIC, 1996, p. 44), que possibilitou o emprego de linguagens computacionais que buscam estruturar e padronizar a forma como as informações

devem ser representadas, resultando na criação de estruturas capazes de promover a representação dos mais distintos tipos de conteúdo.

Considerando a grande produção de recursos informacionais “[...] a representação é um mecanismo indispensável para operacionalizar e garantir a eficiência funcional da recuperação através de sistemas e instrumentos de representação da informação” (AGUIAR; KOBASHI, 2013, não paginado).

No entanto, ressalta-se que a eficiência no processo de recuperação depende diretamente da precisão com que as informações são representadas e, o uso de tecnologias com maior capacidade semântica torna-se uma tendência dado ao rápido desenvolvimento e automação do processo de representação da informação na *Web*.

Pereira e Bufrem (2007, p. 22) destacam que, em um cenário em que “[...] a informação e o conhecimento dela resultante são as novas “moedas correntes”, as comunicações instantâneas ocorrem em velocidade sem precedente na história, provocando uma expansão dos seus domínios”.

Do ponto de vista prático, uma representação eficiente torna ágil o processo de busca informacional uma vez que, o resultado dessa atividade visa atender as necessidades específicas de cada usuário no qual, as informações armazenadas em bibliotecas ou banco de dados, são registradas com o propósito de facilitar sua recuperação (HJORLAND, 1992; SILVA, 2013). Ou seja, a importância nesse processo está no fato de proporcionar ao usuário de sistemas informatizados maior autonomia em suas escolhas informacionais.

Dito isso, as linguagens utilizadas para a representação de recursos são “[...] muitas vezes consideradas como linguagens com propósitos semânticos [...] para a adição de dados sintáticos” (MOTA; KOBASHI, 2016, p. 15), permitindo que uma maior gama de informações e dados possa ser interligada através do uso de estruturas interoperáveis.

Nesse contexto, Santos e Sant’Ana (2013, p. 205) definem dado como um “[...] elemento básico nos fluxos informacionais [...] uma unidade de conteúdo necessariamente relacionada a determinado contexto e composta pela tríade entidade, atributo e valor”. Com isso, os dados se “[...] transformam em objetos qualificados com atributos”, no qual se tem a possibilidade de reutilização automatizada da informação e maior capacidade de compartilhamento entre

sistemas (BAX, 2001, p. 34).

Dado ao rápido desenvolvimento da Internet e a criação de novas tecnologias de representação, o processo de disponibilização dos dados tem acarretado uma mudança paradigmática na forma como enxergamos a *Web*, influenciando novas tendências de pesquisa envolvendo *Linked Data*; *Data Science*; Publicação Ampliada e *Web Semântica*, que em essência possuem em comum a disponibilização de dados em formatos legíveis por computadores, possibilitando a interligação dos recursos informacionais e agregação semântica.

As denominadas tecnologias semânticas caracterizam-se como linguagens que possibilitam ir além de representações sintáticas, descrevendo computacionalmente aspectos semânticos que vão além da estrutura interna dos metadados que descrevem os recursos informacionais, entre as quais pode-se destacar o *Resource Description Framework* (RDF).

Desenvolvido em 1997 pelos editores Ora Lassila e Ralph Swick, o RDF foi projetado para ser um mecanismo para trabalhar com metadados que promovem o intercâmbio de informações entre processos automatizados sem perda de significado (ALESSO; SMITH, 2005; FURGERI, 2006; MONTEIRO; JACYNTHO, 2016).

Para Lassila e Swick (1999, tradução livre), o RDF fornece uma melhor interoperabilidade entre aplicativos que trocam informações permitindo que o processamento automatizado dos recursos da *Web* possa ser usado em uma variedade de áreas de aplicação, como por exemplo: na descoberta de recursos para fornecer melhores recursos de mecanismo de pesquisa, na catalogação para descrever as relações entre conteúdo disponíveis em um determinado site.

Para Jacob (2003) e Ramalho (2006), a aplicação do RDF em ambientes digitais possibilita adicionar semântica a um recurso informacional independente de sua estrutura, permitindo a reutilização de metadados estruturados e/ou esquemas de metadados. Ou seja, o uso de metadados possibilita “[...] descrever o recurso ou objeto informacional de modo a permitir sua identificação, localização, recuperação, manipulação e uso” (CAMPOS, 2007, p. 21).

As novas tecnologias de representação têm como objetivo principal permitir que os dados publicados na *Web* possam ser processados de maneira automática, favorecendo agregações semânticas e permitindo que pessoas e sistemas

computacionais possam ter acesso às informações de maneira rápida e precisa (BERNERS-LEE, 2009; CUEVA YERBA; MEDINA BRAVO, 2015; ARAKAKI, 2016).

Com inúmeras tecnologias disponíveis para a representação e intercâmbio de dados, foi desenvolvido um conjunto de boas práticas utilizadas para a publicação e conexão de dados estruturados na *Web*, mais conhecido como *Linked Data*, que permite conectar diversos recursos, independente do domínio (BIZER et al., 2008; BIZER; HEATH; BERNERS-LEE, 2009; HYLAND; ATEMEZING; VILLAZÓN-TERRAZAS, 2014), relacionando os dados que não estavam ligados anteriormente ou usando a *Web* para reduzir as barreiras para interligar dados atualmente vinculados através de outros métodos (BERGMAN, 2017).

O RDF é a tecnologia que possibilita essa realidade ao proporcionar o estabelecimento de relacionamentos formais entre recursos e dados, estabelecendo uma rede de compartilhamento cada vez mais eficiente na rede mundial de computadores.

Nesse sentido, a proposta do *Linked Data* pode ser entendida como uma abordagem para codificar dados no nível mais granular, na qual dados e objetos podem vir de sistemas de vocabulário padrão como o *Dublin Core* e de esquemas de codificação de vocabulário na forma de vocabulários controlados, ontologias, taxonomias, arquivos de autoridade de nomes, sistemas classificatórios e afins (MÉNDEZ; GREENBERG, 2012), conforme representado na figura 1.

Como apresentado na figura 1, a partir do RDF novas tecnologias foram desenvolvidas possibilitando que o movimento dos dados interligados tivesse um aumento expressivo ao incorporar novos modelos e métodos de representação na *Web*, proporcionando um maior compartilhamento de recursos nesse ambiente.

O uso dessas tecnologias propicia inúmeras vantagens, dentre as quais os diversos recursos produzidos para ambientes digitais podem ser descritos em colaboração mútua entre diversas bibliotecas e interligados a dados disponibilizados por outras instituições ou até mesmo indivíduos, permitindo que todos possam contribuir com conhecimentos únicos de forma que as informações sejam reutilizadas e recombinaadas (BAKER et al., 2011, não paginado, tradução livre).

Assim, os metadados utilizados nesse processo são “[...] elementos representativos da informação no ambiente digital que visam proporcionar o acesso a uma grande variedade de informações e multiplicidade de formatos” (DANTAS; CORDULA; ARAÚJO, 2016, p. 51).

Nesta perspectiva os formatos bibliográficos também têm evoluído ao longo do tempo, entre os novos padrões de representação destaca-se o *Bibliographic Framework* (BIBFRAME), um modelo definido na linguagem RDF que fornece uma base para a descrição bibliográfica contemporânea na *Web* ao ser fundamentado na proposta de dados interligados e possibilita a descrição formal dos relacionamentos entre os recursos informacionais, conforme apresentado a seguir.

3 LINGUAGENS DE REPRESENTAÇÃO

O conceito de Linguagem é utilizado tanto para se referir à capacidade humana de aquisição e utilização de sistemas complexos de comunicação, como também para qualquer meio sistemático de comunicação que se utiliza de signos para troca de mensagens.

Com o desenvolvimento das tecnologias surgiu o conceito de Linguagem Computacional que pode ser definida sucintamente como um conjunto de símbolos e regras formais utilizadas para a definição de sentenças que representam e processam instruções para computadores.

O estudo sobre as aplicações das linguagens computacionais sempre foi, desde a criação dos primeiros sistemas digitais, “[...] um tema recorrente e determinante para o avanço das pesquisas, tendo impulsionado a criação de uma grande variedade de linguagens com diferentes características e propósitos” (MARTINS; RAMALHO, 2016, p. 735), que utilizam marcações <tags> para descrever e estruturar recursos informacionais.

O início do movimento de ‘codificação genérica’ é creditado a William Tunnicliffe que em 1967, durante uma reunião no escritório de imprensa do governo do Canadá, fez uma apresentação abordando a separação do conteúdo informacional do formato dos documentos.

Em 1969 um designer de livros chamado Stanley Rice propôs a ideia de um catálogo universal de *tags* de “estrutura editorial” parametrizada. Assim, Norman Scharpf, então diretor da associação norte-americana *Graphic Communications Association* (GCA) reconheceu a importância dessas tendências e estabeleceu um projeto genérico de codificação para a indústria editorial, que culminou no desenvolvimento de um padrão denominado como *GenCode*.

A partir da disseminação do uso dos computadores proliferam as áreas de utilização e conseqüentemente os tipos de documentos, de modo que cada documento possuía requisitos diferentes, exigindo conjuntos específicos de anotações. Tal cenário evidenciou a necessidade de padronizar a forma de inserir, processar e manipular as anotações dos documentos.

Dado à multiplicidade de conteúdos disponíveis em ambientes digitais como: artigos, filmes, fotos, dentre outros, o processo de representação requer a utilização

de tecnologias cada vez mais eficientes, de modo a permitir uma melhor recuperação, pois, individualmente ninguém é capaz de lembrar onde cada fragmento de informação pode ser encontrado e, quais são as suas relações com outros recursos (GNOLI, 2009).

Com isso, a recuperação de informações depende diretamente da forma como os recursos informacionais são representados, cujo êxito dessa atividade necessita o uso de uma linguagem estruturada e padronizada, pois as mesmas permitem aos computadores uma melhor compreensão e interpretação semântica dos termos utilizados nesse processo.

Nos tópicos a seguir, abordar-se-á as principais linguagens usadas nos processos de representação e criação de modelos de dados e, suas contribuições no campo da Ciência da Informação.

3.1 *Standard Generalized Markup Language (SGML)*

Criada em 1969, pelos pesquisadores Charles Goldfarb, Edward Mosher e Raymond Lorie, a *Generalized Markup Language (GML)* foi a primeira Linguagem de Marcação amplamente reconhecida e serviu de modelo estrutural para o desenvolvimento de outras linguagens.

Após um período de pesquisas voltadas para a GML, Goldfarb continuou a pesquisar sobre estruturas de documentos, criação de conceitos adicionais, referências curtas, processamento de *links* e tipos de documentos simultâneos que não faziam parte da proposta original da GML, estendendo gradativamente suas funcionalidades.

Em 1978 o comitê da *American National Standards Institute (ANSI)* iniciou um projeto para o desenvolvimento de uma linguagem padrão para descrição de textos baseadas na GML. A partir desses estudos foi desenvolvida uma metalinguagem, ou seja, uma linguagem capaz de descrever e definir outras linguagens de marcação denominada de *Standard Generalized Markup Language (SGML)*, dando origem em 1986 ao padrão ISO 8879/86 (EDWARDS, 1997; SENGUPTA; DILLON, 1997; MILLER JR, 2003).

Para Brüggemann-Klein e Wood (1997, tradução livre) a criação do padrão SGML disponibiliza uma metalinguagem sintática para a definição de sistemas de

marcação textual, na qual cada sistema de marcação é especificado, essencialmente, por uma gramática livre de contexto.

Amplamente difundida, a linguagem SGML logo passou a ser utilizada em grande escala em projetos governamentais e comerciais para o compartilhamento de documentos que pudessem ser legíveis por sistemas computacionais. Com tais características, uma das grandes vantagens da criação da SGML, à época, era a possibilidade de descrever documentos a partir de sua própria gramática, permitindo informar através de um conjunto de *tags* o que cada documento representava (ZSILAVECZ, 1993; TEIXEIRA, 2002; MORENO; BRASCHER, 2007).

A SGML é uma linguagem de marcação que pode ser usada por diferentes sistemas operacionais informatizados uma vez que tem código aberto, podendo o usuário criar documentos, independente da plataforma e do sistema na qual seja utilizada (SENGUPTA; DILLON, 1997).

Almeida (2002, p. 6) destaca que, um dos principais objetivos com a criação da SGML “[...] é garantir que documentos codificados de acordo com suas regras possam ser transportados de um ambiente de *hardware* e *software* para outro, sem perda de informação”.

Cada documento da SGML carrega consigo sua própria especificação formal, o *Document Type Definition* (DTD), que possibilita aos sistemas computacionais interpretarem corretamente as marcações existentes. Com isso, uma DTD determina como um documento deve ser representado além da forma como deve ser estruturado, sendo considerada,

[...] um conjunto de regras que define as instruções que serão enviadas ao analisador sintático para o documento que está para ser analisado, podendo incluir um conjunto de declarações de elementos e de atributos e as entidades, notações e comentários que você quer usar (TESCH JR, 2002, p. 19).

No campo da Ciência da Informação, as aplicações da linguagem SGML deram-se pela possibilidade de representação de documentos de maneira estruturada a partir da utilização de marcações adotadas em uma estrutura interpretável por computadores e uso compatível com diferentes tipos de sistemas operacionais disponíveis à época.

Com o aumento na criação de documentos para os ambientes digitais, tal recurso possibilitou à representação documental de forma padronizada, além de

permitir um armazenamento mais amplo de conteúdos informacionais e uma melhor recuperação desses recursos.

Contudo, dada a complexidade de utilização da linguagem SGML, de modo geral, sua aplicação não favorecia a criação de nichos profissionais especializados em aplicações específicas para esse padrão. Sobretudo, é uma linguagem amplamente utilizada em instituições governamentais e de setores de manufatura como o da fabricação de aeronaves, por exemplo.

Devido ao desenvolvimento exponencial da internet, no final da década de 1980 o pesquisador Tim Berners-Lee, do Conselho Europeu para a Pesquisa Nuclear (CERN), propôs uma aplicação da SGML incorporando o uso de *hiperlinks* para representar documentos hipertextuais no ambiente *Web*, dando origem a *HyperText Markup Language* (HTML), abordada no subtópico a seguir.

3.2 *HyperText Markup Language* (HTML)

A *HyperText Markup Language* (HTML) é uma linguagem baseada em SGML e foi a primeira linguagem de marcação que passou a ser amplamente conhecida pelo público em geral.

Desenvolvida, principalmente, como uma linguagem semântica para descrever documentos científicos, HTML pode ser usada, também, para descrever uma série de outros tipos de documentos já que é uma linguagem não proprietária e pode ser criada e processada por uma ampla gama de ferramentas a partir de editores de texto simples.

Caracterizando-se como uma aplicação da SGML, a linguagem HTML usa um conjunto de marcações pré-definidas em sua DTD para representar os diferentes tipos de conteúdos que podem ser apresentados em uma página da *Web*. Com tais características, a HTML logo se torna uma das linguagens mais utilizadas mundialmente e isso contribuiu diretamente para o surgimento de uma enorme quantidade e variedade de documentos para a *Web*, possibilitando assim a disseminação de informações.

Marcondes (2007) destaca que, uma das principais características da linguagem HTML é sua estrutura flexível e de fácil manuseio, estabelecendo apenas o que é título texto, subtítulo, local das imagens, sendo considerada até hoje, a

linguagem de marcação mais simples e fácil de ser manipulada.

Para Furgeri (2006), ao se elaborar um documento em HTML deve-se sempre utilizar no título termos que diz respeito ao documento representado, pois um título bem elaborado contribui diretamente para uma recuperação de documentos eficaz, uma vez que, na maioria das vezes ele será exibido nos resultados de mecanismos de busca utilizados na internet.

Na figura 2 é apresentado um exemplo de um documento simples em HTML. Os documentos HTML consistem em uma árvore de elementos e texto e, cada elemento é denotado na fonte por uma *tag* de início, "<body>", e uma *tag* final, "</body>".

Figura 2 - Exemplo de documento HTML

```
<!DOCTYPE html>
<html>
  <head>
    <title>Exemplo</title>
  </head>
  <body>
    <h1>Exemplo</h1>
    <p>Esse é um <a href="demo.html">simples</a> exemplo.</p>
    <!-- comentários -->
  </body>
</html>
```

Fonte: Elaborado pelo autor

As primeiras versões da HTML foram definidas com regras sintáticas flexíveis, cujo objetivo era favorecer uma maior popularização da *Web* possibilitando representar informações de diversas aplicações. Desenvolvido com foco para a criação de páginas da *Web*, a linguagem HTML tem as seguintes prerrogativas:

- descrever a estrutura de páginas da *Web* usando as marcações;
- seus elementos são a base de construção de páginas HTML;
- seus elementos são representados por *tags*.

Desde sua criação, algumas versões foram publicadas com o intuito de melhorar seu desempenho. São elas: HTML+, HTML 2.0 e HTML 3.0, na qual foram propostas algumas mudanças para enriquecer as possibilidades da linguagem. Contudo, apenas em 1997 o *World Wide Web Consortium (W3C)* publicou a versão

HTML 3.2, definindo-a como uma recomendação oficial.

Devido ao rápido crescimento da *Web* e o desenvolvimento de novas ferramentas computacionais, a HTML logo foi atualizada para a versão 4.0, em 1999, e no ano seguinte, 2000, é atualizada para incorporar os recursos da *eXtensible Markup Language* (XML) em sua estrutura, dando origem a *eXtensible HyperText Markup Language*¹ (XHTML).

No quadro 1 são destacadas as principais atualizações ocorridas na HTML até a última versão disponível.

Quadro 1 - Versões da linguagem HTML

Versão HTML	Data de criação	Alterações	Observações
HTML 1.0	1992	Versão inicial. Linguagem de marcação usada pela <i>World Wide Web</i> para popularização da <i>Web</i> .	Limitada a DTD da SGML, desenvolvida por Berners-Lee.
HTML 2.0	Nov./1995	Adicionado representação de imagem com a tag .	Uma iniciativa desenvolvida para o Mosaic Browser.
HTML 3.0	Não implementado	Tenta-se incluir <i>tags</i> adicionais: <background>, <frame>, , <blink>, <marquee>, <iframe>, <bgsound>entre outros.	Uma iniciativa da W3C para incorporar recursos, sendo introduzidos independentemente pela Netscape e a Microsoft.
HTML 3.2	Jan. 1997	A versão 3.2 atenuou as alterações propostas para a versão 3.0 para ser coerente com as normas existentes.	Estabeleceu uma base comum para provedores de conteúdo e desenvolvedores de navegador.
HTML 4.0	Dez. 1997	Amplia o HTML com mecanismos para folhas de estilo, scripts, quadros, objetos de incorporação, oferecendo melhor acessibilidade para pessoas com deficiências.	Permite uma indexação mais eficaz de documentos para mecanismos de pesquisa, tipografia de melhor qualidade, melhor conversão de texto em fala, melhor hifenização etc.
HTML 4.01	Dez. 1999	HTML 4.01 é uma revisão da Recomendação do HTML 4.0 lançada pela primeira vez em dezembro de 1997.	A especificação XHTML 1.0 depende do HTML 4.01 para os significados dos elementos e atributos XHTML. Isso permitiu

¹ Disponível em: <https://www.w3.org/MarkUp/#recommendations>

			reduzir consideravelmente o tamanho da especificação XHTML 1.0.
XHTML 1.0	2000	Reformulação da HTML 4.0 como uma aplicação XML.	Os tipos de documentos da Família XHTML são todos baseados em XML e foram projetados para trabalhar em conjunto com agentes de usuários baseados em XML.
XHTML 1.1	Nov. 2010	Inclui uma implementação de idioma do XML <i>Schema</i> e integra o atributo <i>lang</i> para aumentar a compatibilidade com os usuários e as Tecnologias Assistivas.	Segunda edição disponível em: https://www.w3.org/TR/xhtml11/xhtml11-rec-diff.html
XHTML 2.0	Dez. 2010	É composta de um conjunto de módulos XHTML que, juntos descrevem os elementos e atributos da linguagem e seu modelo de conteúdo. Atualiza muitos dos módulos definidos na XHTML <i>Modularization</i> 1.1 [XHTMLMOD] e inclui as versões atualizadas de todos esses módulos e suas semânticas.	Foi projetado para corrigir erros e insuficiências no campo do autor nas versões de HTML e XHTML 1. Usa módulos do XHTML Access [XHTMLACCESS], da Função XHTML [XHTMLROLE], da Sintaxe RDFa [RDFASYNTAX], do Ruby [RUBY], dos Eventos XML [XMLEVENTS] e dos XForms [XFORMS].
HTML 5.2	Dez. 2017	Destina-se a autores de documentos e scripts, implementadores de ferramentas que operam em páginas que usam os recursos definidos nesta especificação e indivíduos que desejam estabelecer a correção de documentos ou implementações com relação à requisitos desta especificação.	Essa especificação é limitada a fornecer uma linguagem de marcação de nível semântico e APIs de script de nível semântico associadas para a criação de páginas acessíveis na Web, desde documentos estáticos a aplicativos dinâmicos.

Fonte: Elaborado pelo autor

A linguagem XHTML, lançada em 2000, é uma família de tipos de documentos atuais e futuros disponíveis na *Web*, cujos documentos são todos baseados em XML e, em última instância, são projetados para trabalhar em conjunto com agentes de usuário baseados em XML (PEMBERTON, 2013, tradução livre). Contudo, essa versão não substitui a linguagem HTML e tampouco a linguagem

XML, que tem propósitos distintos entre si.

A XHTML 1.0 foi a primeira grande mudança da HTML, desde a versão HTML 4.0, e traz o rigor da XML, uma linguagem mais complexa e completa, para páginas da *Web*, sendo considerada a peça chave do trabalho do W3C para criar padrões que forneçam páginas da *Web* mais ricas em uma gama cada vez maior de plataformas de navegadores, incluindo telefones celulares, televisões, carros, etc (ALESSO; SMITH, 2005; PEMBERTON, 2013).

Pemberton (2013) destaca que, a XHTML 1.0 possui três variantes e possibilita ao usuário especificar qual delas utilizar ao desenvolver um documento em ambientes digitais. São elas:

- XHTML 1.0 *Strict* – é usado em conjunto com a linguagem W3C's *CascadingStyleSheet* para se ter fonte, cor e efeitos de *layout* para uma marcação estrutural limpa e livre de qualquer marcação associada ao *layout*;
- XHTML 1.0 *Transitional* – é usado para gerar páginas da *Web* para o público em geral;
- XHTML 1.0 *Frameset* – é usado para particionar à janela do navegador em duas ou mais *frames*.

Já na versão XHTML 2.0, muitos dos módulos que descrevem os elementos e atributos definidos na modularização da XHTML são atualizados e incorporam as versões atualizadas de todos esses módulos e sua semântica. Essencialmente essa versão consiste em um pacote de várias partes utilizadas na *Web*.

Para Almeida (2002), o uso da HTML para a representação de páginas *Web* possibilitou que os dados fossem apresentados em uma estrutura simples e de fácil leitura, sendo que cada versão da HTML se fundamenta em um novo DTD, limitando-se às marcações previamente definidas.

Contudo, sua concepção baseada em marcações fixas, conhecidas como *tags*, trouxe limitações para a representação de informações uma vez que, sendo uma linguagem de marcação voltada, principalmente, para a apresentação de documentos em ambientes digitais, a HTML apenas define a forma como a informação é exibida não se preocupando com o significado da palavra e, ocasionalmente é usada para lidar com dados.

Alesso e Smith (2005, p. 62, tradução livre) destacam que, com o crescente aumento no número de documentos e conteúdos disponíveis na internet, a HTML “[...] não suporta a necessidade de estrutura e semântica para aplicações complexas exigidas pela futura arquitetura da *Web*”.

Com o intuito de promover uma melhor funcionalidade da linguagem, desde a incorporação de recursos da linguagem XML, o grupo de trabalho do W3C responsável pelo projeto, publicou em 2014 a especificação HTML5. Essa versão representa um marco no desenvolvimento da HTML, pois apresenta a totalidade da tecnologia aplicada na linguagem HTML (HICKSON et al., 2014; FAULKNER et al., 2017), tendo sua última especificação, HTML 5.2, publicado em 2017.

No campo da Ciência da Informação, a aplicação da linguagem HTML dá-se no fato de permitir a automação dos catálogos de bibliotecas que evoluíram do catálogo manual para o catálogo *on-line*, denominados na CI como *Online Public Access Catalog* (OPAC).

Com a automação dos catálogos de bibliotecas, a disseminação de informações tornou-se um processo rápido e eficiente, uma vez que “[...] os diversos tipos de catálogos são considerados como uma forma de mediação entre os recursos informacionais e seus usuários”. (SIMIONATO; ARAKAKI; SANTOS 2017, p. 450).

Os autores Simionato, Arakaki e Santos (2017) destacam que a descrição dos recursos informacionais presente nos catálogos devem ser estruturados de modo a permitir fácil acesso e localização dos itens representados. Com isso,

[...] é possível pelos módulos de OPAC realizar pesquisas por autor, título e assunto, cumprindo as funções das tradicionais fichas catalográficas, porém com mais rapidez [...], além disso, considera-se que o avanço tecnológico contribui para a otimização da operação de tratamento da informação em bibliotecas, sendo através dessa operação que o bibliotecário executará a função de inserir os dados no catálogo, dos documentos que compõem o acervo (SOUSA; FUJITA, 2012, p. 65).

Segundo Oliveira (2009), a simples busca de informação em um catálogo *on-line* é um processo relativamente rápido e dinâmico, tendo em vista que os documentos representados possuem relações feitas através de *links*, permitindo recuperar resultados de busca de forma mais ampla.

A aplicação da HTML na área da CI permitiu diversos estudos voltados para a

criação e/ou o desenvolvimento de novos meios de representação e disseminação de informações ao utilizar a linguagem para potencializar o compartilhamento de conteúdo entre instituições através dos catálogos *on-line*.

Contudo, a HTML é uma linguagem limitada para os padrões de compartilhamento de dados e informações requeridos no ambiente *Web*, na qual uma das únicas formas de interação que ela dispõe, sem o apoio de outras linguagens, são os *links* entre páginas (ALESSO; SMITH, 2005; MARCONDES, 2007).

Tais limitações semânticas impulsionaram o desenvolvimento de uma nova linguagem de marcação com capacidade de aplicações mais amplas no contexto dos ambientes digitais, denominada de *eXtensible Markup Language* (XML), abordada no tópico a seguir.

3.3 *eXtensible Markup Language* (XML)

Criada pelo engenheiro Jon Bosak, em 1996, a *eXtensible Markup Language* (XML) foi desenvolvida com base na linguagem SGML e teve sua recomendação divulgada em 1998. Diferente da linguagem HTML, que se caracteriza como um esquema de aplicação SGML, a linguagem XML é um subconjunto simplificado de SGML e permite a formalização de outras linguagens.

Considerada uma metalinguagem, a XML possibilita a criação de novas linguagens e a definição de marcações personalizadas por meio da criação de seus próprios DTDs, permitindo que elementos de marcação sejam definidos pelo seu desenvolvedor, além de ser uma linguagem projetada para o armazenamento, exportação e compartilhamento de dados (BAX, 2001; HEITLINGER, 2001; RAMALHO, 2006).

Para Almeida (2002, p. 7), a especificação do XML “[...] define um dialeto simples do SGML, permitindo o processamento dos documentos na Internet e utilizando-se de recursos inexistentes no HTML”.

Inicialmente, um dos fatores que dificultaram a disseminação da linguagem XML foi uma concepção equivocada de que a mesma fosse uma linguagem sucessora da HTML (RAMALHO, 2006). Com propósitos distintos, HTML tem como foco principal a representação dos documentos como páginas *Web*, enquanto que

XML busca estruturar formalmente os documentos para que estes sejam legíveis por sistemas computacionais e possibilite uma melhor semântica dos dados representados.

Segundo Bax (2001) e Furgeri (2006), uma das funções básicas da linguagem XML é se ocupar com a representação de informações de forma estruturada e não definir como essas informações devem ser apresentadas. Assim, a função de apresentação de conteúdo é uma das especificidades da linguagem HTML, cujas marcações definem a aparência dos dados e a posição dos títulos e, em XML as marcações definem a estrutura e o significado dos dados e o que eles representam.

A linguagem XML desempenha um papel semelhante ao de uma linguagem universal podendo ser utilizada para o desenvolvimento de diferentes linguagens de uso especial, além de ser conhecida pela simplificação das comunicações entre plataformas não homogêneas, permitindo um melhor relacionamento entre os sistemas (NEEDLEMAN, 1999; DAUM; MERTEN, 2002; OPARA; SRIVASTAVA, 2003).

Em ambientes digitais, a linguagem XML “[...] pode se tornar qualquer coisa que um documento necessite que ele seja para distribuir informações na *Web* ou entre aplicativos (*software*) sem as restrições e limitações de HTML” (MILLER JR., 2003, p. 11). Ou seja, XML estabelece uma padronização no processo de representação de informações, uma vez que os requisitos de HTML não atendem a interoperabilidade informacional na *Web*.

Os documentos em XML dividem-se em duas grandes categorias: os centrados em dados e os centrados em documentos. Os centrados em dados são aqueles em que a XML é utilizado para transporte de dados (ordem de vendas, registro de pacientes, etc.) e, os centrados em documentos são aqueles em que a XML é usado para refletir a estrutura de classes particulares de documentos, como livros com capítulos, manuais de usuários e *feeds* de notícias (LIBRARY OF CONGRESS, 2015, tradução livre).

A partir de 2001, além da definição de tipos de dados por meio de DTDs, a XML também passou a suportar o XML *Schema Definition* (XSD), ou simplesmente XML *Schema*, desenvolvido com o objetivo de superar as limitações existentes no DTD e facilitar a codificação a partir da sintaxe XML.

O XML *Schema* é uma tecnologia projetada para substituir os DTDs com uma

abordagem mais poderosa e intuitiva e, permite que as linguagens de marcação descrevam cuidadosamente os elementos e atributos que um documento em XML deve usar (DESMARAIS, 2000; ALESSO; SMITH, 2005). Contudo, devido à similaridade de funções, em casos específicos, o termo *XML Schema* é utilizado para evitar qualquer equívoco que possa existir com o termo DTD (DESMARAIS, 2000).

Assim, uma DTD determina qual parte de um documento pode ou não ser utilizada durante sua elaboração, bem como descreve o conteúdo de um documento em XML.

Na figura 3 é apresentado um exemplo básico de um documento em XML e sua descrição nas marcações pré-estabelecidas da estrutura, usada, nesse exemplo, para a criação de um catálogo simples.

Figura 3 - Exemplo básico de um catálogo em XML

```
<?xml version="1.0"?>
<catalogo xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="catalogo.xsd">

  <livro>
    <titulo>Caixa-Preta</titulo>
    <autor>Ivan Sant'Anna</autor>
    <descricao>Relatos de desastres aéreos brasileiros</descricao>
    <ISBN>978-85-390-0180-4</ISBN>
  </livro>

</catalogo>
```

Fonte: Elaborado pelo autor

Segundo o W3C, publicado por Quin (2015, tradução livre), o *XML Schema* é utilizado para expressar restrições sobre documentos em XML e pode ser usado para:

- fornecer uma lista de elementos e atributos em um vocabulário;
- associar tipos, como *integer*, *string*, etc., ou mais especificamente como um *hatsize*, *sock_colour*, etc., com valores encontrados em documentos;
- restringir onde os elementos e atributos podem aparecer, e o que aparecer dentro desses elementos;

- fornecer documentação que seja legível por humanos e processáveis por computadores;
- dar uma descrição formal de um ou mais documentos.

Tanto uma DTD quanto o XML *Schema* podem utilizar, como formatos de dados, as relações estruturais entre os elementos e os vocabulários em XML que identificam os elementos utilizados em formatos de dados específicos e, somente os dados em XML válidos exigem uma DTD.

A estrutura do XML *Schema* é definida a partir do elemento raiz `<xsd:element>`, de modo que os elementos podem ser definidos como elementos simples ou complexos (DAUM; MERTEN, 2002). Assim, XML *Schema* permite “[...] definir linguagens de marcação descrevendo cuidadosamente os elementos e atributos que um documento em XML deve usar” (ALESSO; SMITH, 2005, p. 76, tradução livre).

Na figura 4 é apresentado um fragmento do XML *Schema* que descreve formalmente a estrutura e os elementos que compõem um documento em XML, cujo documento XML *Schema* é representado pela declaração ou *namespace*/URI de referência `<xsd:schemaxmlns:xsd="http://www.w3.org/2001/XMLSchema">` e, é iniciado com uma declaração XML `<?xml version="1.0"?>`, que é a versão utilizada para formalização dos esquemas.

Figura 4 - Fragmento de XML Schema

```

<?xml version="1.0"?>
- <xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <xsd:element type="TipoCatalogo" name="catalogo"/>
  - <xsd:complexType name="TipoCatalogo">
    - <xsd:sequence maxOccurs="unbounded" minOccurs="0">
      <xsd:element type="TipoLivro" name="livro"/>
    </xsd:sequence>
  </xsd:complexType>
  - <xsd:complexType name="TipoLivro">
    - <xsd:sequence>
      <xsd:element type="xsd:string" name="titulo"/>
      <xsd:element type="xsd:string" name="autor"/>
      <xsd:element type="xsd:string" name="descricao"/>
      <xsd:element type="xsd:string" name="ISBN"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:schema>

```

Fonte: Elaborado pelo autor.

Os seguintes elementos compõem um documento em XML *Schema*, no qual o elemento `<xsd:schema>` contém a definição completa do esquema além de definir o *namespace*.

- `<xsd:element>` - define um elemento raiz dentro de um documento;
- `<xsd:complexType>` - define quais atributos esses elementos possuem e quais elementos filho contém;
- `<xsd:simpleType>` - define o tipo de um elemento folha ou de um atributo (MICROSOFT, 2015, tradução livre).

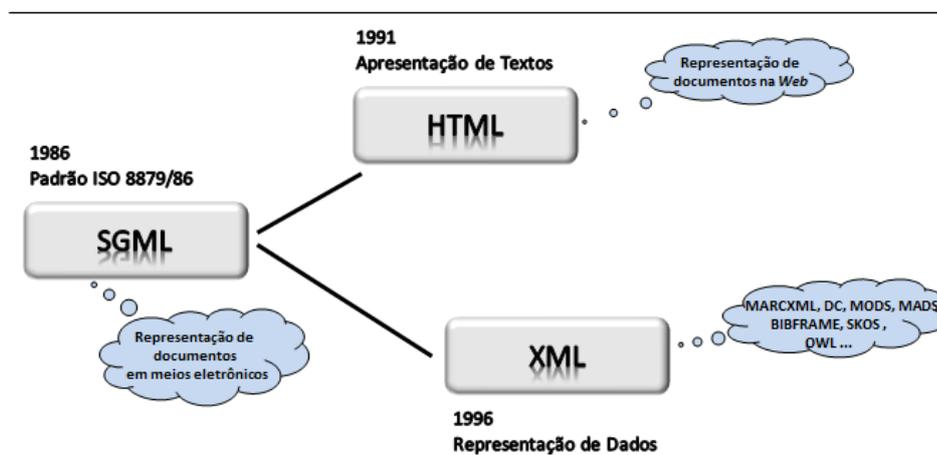
Os elementos de um esquema XML são agrupados por suas seguintes funções: elementos de nível superior, partículas, vários documentos em XML e *namespace*, restrições de identidade, atributos, atributos nomeados, definições de tipos simples e complexos.

Para Ramalho (2006, p. 64), a diferença básica entre elementos simples e complexos é que os elementos simples não contêm subelementos como os elementos complexos e, os atributos “[...] são fontes de informações adicionais sobre um elemento e sempre possuem um valor que deve ser delimitado por aspas”.

Dado as inúmeras vantagens de manipulação, a linguagem XML possibilita maior flexibilidade e liberdade para a representação a partir da definição de marcações personalizadas extensíveis, contribuindo para uma melhor estruturação e enriquecimento dos dados representados, o que a torna uma escolha lógica para o trabalho de comunicação padronizada em ambientes digitais e *Web* (ALESSO; SMITH, 2005).

Na figura 5 é apresentado um extrato da evolução das linguagens de marcação a partir da SGML até a metalinguagem XML, que possibilita desenvolver novas linguagens e padrões de metadados empregados na representação de recursos, como: MARCXML, *Dublin Core*, BIBFRAME, dentre outros.

Figura 5 - Cronologia de desenvolvimento das LMs



Fonte: Elaborado pelo autor

Com uma ampla usabilidade nos processos de representação as aplicações em XML, na área de Ciência da Informação, vão além das concebidas pela linguagem HTML, uma vez que o uso da linguagem XML permite a criação de estruturas com maior capacidade semântica e possibilita o compartilhamento de conteúdo entre diversas instituições de forma eficaz.

A partir da XML, diversas aplicações podem ser desenvolvidas para ambientes digitais, como a criação de ontologias destinadas a propósitos distintos, permitindo que qualquer tipo de documento desenvolvido nessa linguagem seja armazenado de maneira estruturada e recuperado de forma rápida (FURGERI, 2006).

Na área de Ciência da Informação, a aplicação da XML favorece diversos estudos com o objetivo de propor novos métodos e modelos de armazenamento, representação, reuso e compartilhamento de informações e dados permitindo ao usuário a recuperação de conteúdos que atendam suas necessidades informacionais.

Contudo, o crescimento de conteúdo para a Internet é algo expressivo e contínuo e, a linguagem XML, mesmo com todas suas vantagens, apresenta limitações dado ao grande número de conceitos adotados no contexto atual tanto para indexação e representação, quanto para os termos utilizados nos mecanismos de busca. Nesse cenário, novos modelos de representação e organização de informação e dados estão sendo desenvolvidos para suprir tal demanda e proporcionar uma maior capacidade semântica dos termos adotados no ambiente

Web.

Nesse cenário, destaca-se o *Resource Description Framework* (RDF) que é um modelo de que adota a estrutura da linguagem XML para prover um aumento na capacidade de intercâmbio de dados ao vincular relacionamentos existentes entre os diversos documentos disponíveis na *Web* e, que está diretamente inserida nos propósitos das tecnologias semânticas utilizadas para a representação de informações a partir de dados interligados.

3.4 *Resource Description Framework* (RDF)

Desenvolvida em 1997, o RDF foi projetado como uma estrutura para representar informações na *Web* ao trabalhar com metadados que promovem o intercâmbio de informações entre processos automatizados por meio de uma semântica mais abrangente dos recursos representados em domínios distintos (ALESSO; SMITH, 2005; FURGERI, 2006; LASSILA; SWICK, 1999).

A partir da criação de modelos de dados com grande capacidade semântica, o ambiente *Web* constitui-se como uma das mais ricas fontes de informação e com isso há um esforço no sentido de converter a *Web* em uma *Web Semântica* na qual o RDF tornou-se a linguagem escolhida para modelagem, interconexão e fusão de dados (BISCHOF et al., 2011; RAMALHO, 2015).

Cada vez mais presentes na *Web*, dados em RDF e aplicações que se utilizam desses dados permitem uma inferência de conteúdo informacional com maior eficiência na interoperabilidade de informações, uma vez que permite estabelecer relações formais entre informações disponíveis em diferentes domínios (LASSILA; SWICK, 1999; KONTOPOULOS; KRAVARI; BASSILIADES, 2007; BISCHOF et al., 2011).

O RDF é uma aplicação da linguagem XML que se propõe ser uma base para o processamento de metadados na *Web*. Sua padronização estabelece um modelo de dados e sintaxe para codificar, representar e transmitir metadados, com o objetivo de torná-los processáveis por máquina, promovendo a integração dos sistemas de informação disponíveis na *Web* (LASSILA; SWICK, 1999, não paginado, tradução livre).

Considerada a primeira linguagem que busca ir além das representações sintáticas, RDF permite a formalização de relacionamentos agregando informações

semânticas e contextuais que fundamentam a representação semântica entre diversos recursos informacionais (MARTINS; RAMALHO, 2017), promovendo a ampla comunicação entre diversos sistemas na internet.

Para Grácio (2002, p. 50), o RDF é a “[...] base para o processamento de metadados e tem como principal objetivo proporcionar interoperabilidade entre aplicações que trocam informações eletrônicas da Web”. Ou seja, RDF é uma estrutura utilizada para representar e organizar informações na *Web*, cuja sintaxe define um modelo de dados capaz de ligar todas as linguagens baseadas em RDF e suas especificações.

Segundo a recomendação do W3C, publicada por Cyganiak, Wood e Lanthaler (2014), a estrutura principal da sintaxe do RDF é um conjunto de triplas, cada uma consistindo de um Sujeito, um Predicado e um Objeto, no qual o conjunto dessas triplas é chamado de grafo RDF favorecendo a interligação de dados para os mais diferentes propósitos. Assim, o sujeito é um recurso no qual uma sentença está se referindo; o predicado descreve uma característica, propriedade ou algum relacionamento utilizado para descrever algo sobre este recurso e, o objeto é o valor de uma determinada característica do recurso referenciado, podendo inclusive ser outro recurso.

O RDF pode ser usado para inúmeras aplicações em diferentes contextos e propósitos, dentre elas:

- adicionar informações legíveis por sistemas computacionais para uso em *Web pages*. Por exemplo, o uso do vocabulário popular *schema.org*, permitindo que as informações sejam exibidas em um formato aprimorado nos mecanismos de pesquisa ou processadas automaticamente por aplicativos de terceiros;
- enriquecer as bases de dados vinculando-o a uma outra base de dados de terceiros. Por exemplo, um conjunto de dados sobre pinturas poderia ser enriquecido por vinculá-los aos artistas correspondentes no *Wikidata*, permitindo o acesso a uma ampla gama de informações sobre eles e recursos relacionados;
- interligar os *Application Programming Interface (API) feeds*, certificando-se de que os clientes podem facilmente descobrir como acessar mais informações;

- usar as bases de dados publicados como *Linked Data* [LINKED-DATA], por exemplo, para criar agregações de dados em torno de tópicos específicos;
- criar redes sociais distribuídas através da interligação de descrições RDF de pessoas em vários sites da *Web*;
- fornecer uma maneira compatível entre padrões para o intercâmbio de dados entre bases de dados;
- interligar várias bases de dados dentro de uma organização, permitindo que consultas em bases de dados sejam realizadas usando SPARQL (SCHREIBER; RAIMOND, 2014, tradução livre).

O RDF tem por objetivo “[...] definir um mecanismo para descrever recursos que não faz suposições sobre um domínio de aplicação específico, nem define (a priori) a semântica de qualquer domínio de aplicação” (LASSILA; SWICK, 1999, não paginado), permitindo descrever declarações a respeito de recursos sem a necessidade de modificações prévias.

Com isso, o RDF é:

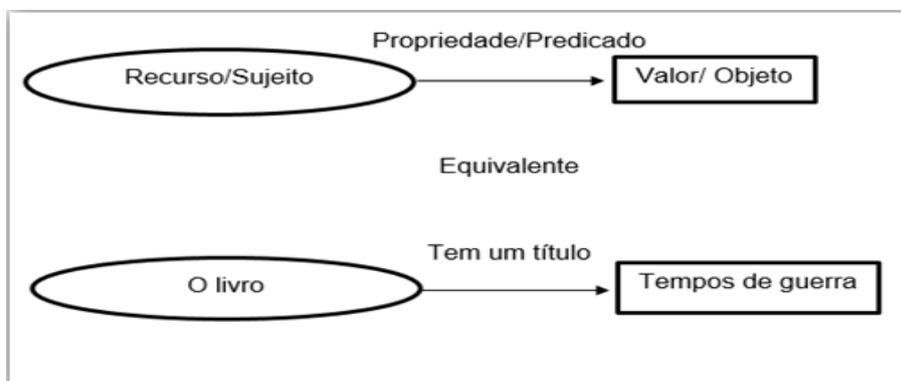
[...] uma base de processamento de metadados que promove a interoperabilidade entre diversas aplicações que variam entre páginas Web ou seu conteúdo, até documentos armazenados em um computador pessoal. É uma linguagem e uma gramática para definir a arquitetura da metainformação na Web, que permite a descrição de recursos de maneira estruturada (PEREIRA; RIBEIRO JR.; NEVES, 2005, p.21).

Dessa maneira, o RDF é um *framework* projetado para utilizar a infraestrutura da WWW aproveitando o armazenamento de informações, a capacidade de comunicação e de processamento de metadados por computadores. Com isso, fornece mecanismos que possibilitam “[...] descrever recursos e seus relacionamentos de modo independente de qualquer implementação e sintaxe, a partir de uma semântica simplificada que pode ser representada utilizando-se a linguagem XML” (RAMALHO, 2006, p. 74).

Enquanto XML possibilita descrever formalmente a estrutura dos documentos, as marcações RDF permitem relacionar formalmente diferentes documentos, contextualizando as relações existentes facilitando a interoperabilidade entre sistemas heterogêneos (GIBBINS; SHADBOLT, 2009).

A figura 6 apresenta uma declaração RDF a partir de um grafo rotulado, denominado de “diagrama de nós e arcos”.

Figura 6 - Exemplo básico de um Grafo da declaração RDF



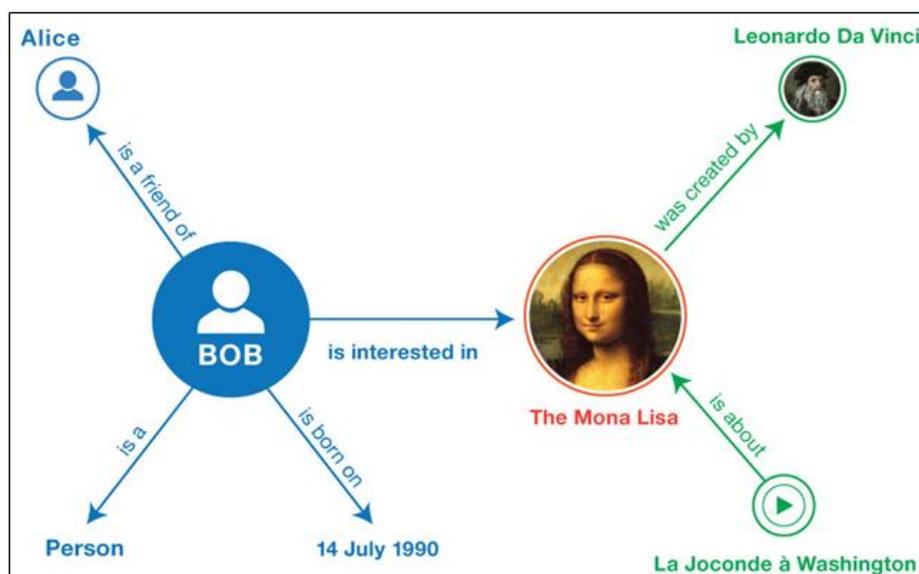
Fonte: Elaborado pelo autor.

O RDF pode ser considerado como uma linguagem para representar diversas informações na *Web* a partir do estabelecimento de ligações formais entre diferentes recursos, independente do tipo de documento representado (BRICKLEY; GUHA, 2004; KLYNE; CARROLL, 2004; MANOLA; MILLER, 2004).

Na figura 7 é ilustrado um exemplo de Grafo RDF de triplas simples. No grafo, Bob é o sujeito de quatro triplas e Mona Lisa é o sujeito de um e o objeto de duas triplas, representado pela seguinte estrutura:

<Bob>	<is a>	<person>
<Bob>	<is a friend of>	<Alice>
<Bob>	<is born on>	<the 4th of July 1990>
<Bob>	<is interested in>	<the Mona Lisa>
<the Mona Lisa>	<was created by>	<Leonardo da Vinci>
<the video 'La Joconde à Washington'>	<is about>	<the Mona Lisa>

Figura 7 - Grafo RDF de triplas simples



Fonte: Schreiber e Raimond (2014).

Schreiber e Raimond (2014, tradução livre) destacam que, essa capacidade de ter o mesmo recurso na posição de sujeito de uma tripla e a posição de objeto de outro, torna possível encontrar conexões entre triplas que é uma parte importante do poder do RDF. Assim, a criação da *Web* de dados é constituída com a descrição de documentos através das relações existentes entre os links e as coisas arbitrárias descritas em RDF, no qual o *Uniform Resource Identifier* (URI) identifica qualquer tipo de objeto ou conceito descrito em uma linguagem compatível com o RDF.

O RDF explora a infraestrutura da *Web* usando uma forma especial de URI e o *International Resource Identifier* (IRI) identifica um recurso e, o *Uniform Resource Locator* (URL) é uma forma de IRI que é utilizado para nomear endereços da *Web* (JAIN, 2012). Com isso, os recursos e os predicados em RDF identificados em URI, possibilitam uma maneira global e única de nomear itens na *Web* (FERREIRA; SANTOS, 2013).

O IRI em um vocabulário RDF geralmente começa com uma “*substring*” comum conhecido como IRI de *namespace* e, alguns *namespaces* do IRI são associados por convenção com um nome curto conhecido como um prefixo de *namespace*, como por exemplo: os prefixos “rdf”, “rdfs” e “xsd” possuem o IRI <http://www.w3.org/1999/02/22-rdf-syntax-ns#>, <http://www.w3.org/2000/01/rdf-schema#> e <http://www.w3.org/2002/XMLSchema#>, respectivamente.

A estrutura de RDF possui duas linguagens de ontologia, o RDF *Schema* e a

Web Ontology Language (OWL). Considerada uma linguagem mais simples, *RDF Schema* é uma extensão semântica que fornece mecanismos para descrever grupos de recursos relacionados e as relações existentes entre esses recursos através de um vocabulário de modelagem de dados para dados em RDF, definindo classes, propriedades e certas restrições globais sobre essas propriedades (BRICKLEY; GUHA, 2014). Já a linguagem OWL é considerada mais expressiva e complexa e, fornece um conjunto maior de construções de modelagem que permite a um designer de ontologias especificar mais ricamente as classes em termos das condições necessárias e suficientes para a adesão a essas classes.

Desenvolvidos em conjunto pelo W3C, o RDF e o RDFS formam um conjunto de ferramentas utilizadas para descrever um recurso, definir um esquema de metadados ou uma ontologia em ambientes digitais. O objetivo do *RDF Schema* é fornecer um vocabulário em XML que possa expressar classes e seus relacionamentos, bem como definir propriedades e associá-las às classes existentes (ALESSO; SMITH, 2005, tradução livre).

O RDFS complementa e estende o RDF definindo uma linguagem declarativa, processável em máquina - uma "metaontologia" ou vocabulário básico de elementos - que pode ser usado para descrever formalmente uma ontologia ou esquema de metadados como um conjunto de classes (tipos de recursos) e suas propriedades (JACOB, 2003, não paginado, tradução livre).

O sistema de propriedades e classes do *RDF Schema* é semelhante aos sistemas de tipos de linguagens de programação orientada a objeto ao descrever as propriedades e classes de recurso a que se aplicam nos mecanismos de domínio e de intervalo descritos na especificação RDF.

Os recursos em RDF podem ser divididos em grupos chamados de classes e, os membros dessas classes são conhecidos como instâncias de classes que são frequentemente identificados por um IRI descritas usando a propriedade "rdf:type" que permite ao usuário indicar que um recurso é uma instância de uma determinada classe (BRICKLEY; GUHA, 2014; GIBBINS; SHADBOLT, 2009).

Apresenta-se na figura 8 um extrato da estrutura de RDF/XML na qual o "rdf:Description" está dentro do "rdf:RDF". Para codificar o grafo em XML, os nós e predicados devem ser representados em termos XML, com os nomes de elementos, atributos, conteúdos de elementos e valores de atributos.

Figura 8 - Estrutura de RDF/XML

```

<?xml:version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.1/"
         xmlns:ex="http://example.org/stuff/1.0/">
  <rdf:Description rdf:about="http://www.w3.org/TR/rdf-syntax-grammar"
                 dc:title="RDF1.1:XML Syntax">
    <ex:editor>
      <rdf:Description ex:fullName="Paulo Martins">
        <ex:homePage rdf:resource="http://purl.org/net/dajobe/" />
      </rdf:Description>
    </ex:editor>
  </rdf:Description>
</rdf:RDF>

```

Fonte: Elaborado pelo autor.

O RDF e a XML são modelos de representação que se complementam, pois se o RDF é um modelo de metadados que define o domínio de endereços por referência e o relacionamento entre si e diversos outros recursos, a XML é a estrutura semântica que torna esses relacionamentos possíveis.

A especificação de RDF é composta de duas partes: a Especificação de Sintaxe, disponível no endereço “<https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>” e, a Especificação de Esquema, disponível no endereço “<http://www.w3.org/TR/2003/WD-rdf-shema-20030123/>”.

A sintaxe de RDF favorece o relacionamento entre si, dos recursos representados em RDF, na qual a linguagem XML é utilizada para maximizar a interoperabilidade de dados. Lassila e Swick (1999, não paginado, tradução livre) destacam que, um dos objetivos do RDF é “[...] tornar possível especificar a semântica de dados baseados em XML de forma padronizada e interoperável”.

Dessa maneira, é considerado um modelo que permite a codificação, troca e reutilização de metadados estruturados, possibilitando a interoperabilidade desses metadados através da criação de mecanismos que suportam convenções comuns de semântica, sintaxe e estrutura (MILLER, 1998; FERREIRA; SANTOS, 2013; RAMALHO, 2015).

Para Miller (1998, tradução livre), devido à grande capacidade de padronizar declarações de vocabulários, o RDF não estipula a semântica para a descrição de recursos, mas fornece a capacidade de definir elementos de metadados específicos às necessidades dos desenvolvedores de sistemas de informação. Com isso, é um

modelo que possibilita criar relacionamentos formais com qualquer parte de um documento, contextualizando as relações existentes (MARTINS; RAMALHO, 2016).

Nessa perspectiva, o RDF desempenha um papel fundamental na oferta de um formato de representação de dados de núcleo para a *Web* semântica, uma vez que as informações podem ser representadas por uma estrutura com recursos em forma de grafo ao descrever metadados sobre recursos WWW como título, criador e data de criação de uma página de blog (GAO, 2011).

Com uma estrutura semântica mais abrangente para os processos de representação e recuperação informacional, o RDF se torna um modelo de grande importância para a área de Ciência da Informação ao proporcionar um compartilhamento de conteúdo de forma mais eficiente, uma vez que possibilita criar vocabulários controlados ao permitir a adição de diversas informações de maneira semântica (FURGERI, 2006).

Uma declaração em RDF pode ser feita de duas formas, sendo que um recurso pode possuir relação com outro recurso ou com um valor literal, como apresentado na figura 9.

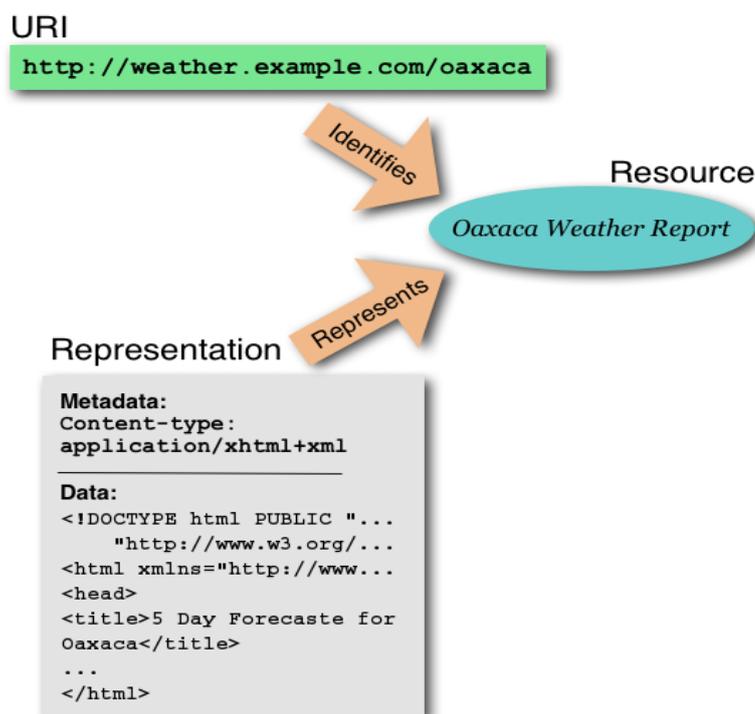
Figura 9 - Relações entre declarações RDF

Relação com outro recurso
<ul style="list-style-type: none"> – URI do recurso URI da propriedade URI do valor <ul style="list-style-type: none"> • Ex: Paulo aluno UFSCar
Relação com um valor literal
<ul style="list-style-type: none"> – URI do recurso URI da propriedade Valor literal <ul style="list-style-type: none"> • Ex: Paulo tem e-mail pgeorgemm22@gmail.com

Fonte: Elaborado pelo autor

Outra forma de representar uma declaração é apresentada na figura 10 e destaca a relação entre o identificador (representado pelo endereço `http://weather.example.com/oaxaca`), o recurso (representado pelo registro “OaxacaWeatherReport”) e a representação (descrito pela aplicação “XHTML+XML”), considerando ainda que o predicado deve ser descrito de maneira formal para que possa ser referenciado por um URI, bem como o sujeito, a fim de que os computadores consigam analisar e compreender o seu significado.

Figura 10 - Declaração RDF



Fonte: Jacobs e Walsh (2004).

O ambiente *Web* possui uma quantidade infinita de informações disponíveis em seu universo e o uso de um modelo como o RDF melhora o acesso a essas informações uma vez que, sua arquitetura é necessária para suportar a descrição de metadados na *Web* (TAUBERER, 2014, tradução livre).

Nessa perspectiva, o RDF fornece meios para publicar tanto um vocabulário legível por humanos quanto por sistemas computacionais, possibilitando o intercâmbio de dados em larga escala, o uso e extensão semântica de metadados entre unidades de informação distintas.

Com a criação das boas práticas de trabalho no âmbito da *Web* para o movimento de dados interligados (MARCONDES, 2012), cujo objetivo é potencializar a interoperabilidade de informações na *Web* a níveis mais elevados, o modelo RDF é a linguagem que melhor representa e formaliza essa realidade.

Ao utilizar os *namespaces* para realizar as inferências de recursos já existentes no ambiente *Web*, o uso do RDF aumenta a diversidade de termos utilizados na representação já que sua estrutura é bastante flexível e sem restrições quanto à utilização de conceitos e, os sistemas computacionais podem distinguir se os relacionamentos existentes, entre recursos, representam um termo ou uma

definição (BECKETT et al., 2014).

Na área de Ciência da Informação, a contribuição de RDF dá-se no uso em aplicações relacionadas à catalogação e à descrição de diversos documentos e informações em ambientes digitais (ALEIXO; SANT'ANA; RAMALHO, 2017), o que contribui para uma disseminação informacional mais eficiente e, permite a recuperação de uma grande variedade de termos pesquisados.

Além das inúmeras aplicações, mencionadas anteriormente, o RDF pode ainda ser usado em conjunto com outros modelos de dados desenvolvidos para potencializar o intercâmbio de informações entre unidades informacionais, permitindo que qualquer tipo de conteúdo seja melhor explorado em ambientes digitais e na *Web* (ALLEMANG; HENDLER, 2008).

Nesse contexto, o desenvolvimento da *Web Semântica* é substancialmente baseado em ontologias que são modelos de dados que representam um dado domínio e são usadas para justificar sobre os objetos desse domínio e as relações existentes entre eles (RAMALHO, 2006; KONTOPOULOS; KRAVARI; BASSILIADES, 2007; PALMÉR, 2012).

Para Berners-Lee (2009, tradução livre), a representação de informações não se trata apenas colocar dados na *Web*, mas sim de fazer links para que uma pessoa ou computador possa explorar de forma mais abrangente a *web* de dados que, com a interligação de dados em ambientes digitais é possível encontrar qualquer outro dado relacionado a partir de um já vinculado previamente.

Assim, a inovação tecnológica desenvolvida ao longo dos anos possibilitou um avanço substancial nos processos de representação de recursos informacionais no ambiente *Web* ao permitir a criação de modelos que potencializam relacionamentos semânticos mais complexos.

3.5 *Web Ontology Language* (OWL)

Desenvolvida pelo W3C, a *Web Ontology Language* (OWL) é uma linguagem da *Web Semântica* projetada para representar um conhecimento rico e complexo, semanticamente, sobre diversos recursos e as relações entre eles e, suportar a inferência entre bases de dados (GAO, 2011; WORLD WIDE WEB CONSORTION, 2012).

No âmbito da *Web* semântica as ontologias em OWL são fundamentais para a criação e desenvolvimento dos processos de inferência informacional já que, necessitam de uma semântica bem definida para potencializar o intercâmbio de dados em ambientes digitais (PASTOR SÁNCHEZ, 2012).

Para Palmér (2012, p. 52, tradução livre), embora a OWL tenha o mesmo objetivo do RDF *Schema*, que é definir classes e propriedades e como elas se relacionam, ela é considerada uma linguagem mais completa e fornece restrições de cardinalidade nas propriedades. Já o RDF *Schema* possibilita definir vocabulários em termos de classes, propriedades e seus domínios e intervalos, sendo o mesmo usado para modelar um domínio específico e editar instâncias de classe (PALMÉR, 2012).

Segundo Gao (2011, p. 20, tradução livre) a semântica definida por RDFS e OWL “[...] pode ser usada para aplicações dentro de ontologias e bases de dados [...] para padronizar e descrever as regras”.

O RDF *Schema* e a OWL fornecem a base para que os recursos possam ser descritos e relacionados entre os diversos domínios, no qual os vocabulários são expressos em RDF, usando termos de RDFS e OWL, que fornecem diversos graus de expressividade em domínios de modelagem de interesse do desenvolvedor (BRICKLEY; GUHA, 2004; BIZER; HEATH; BERNERS-LEE, 2009).

Considerada mais expressiva e complexa para a criação de ontologias, a OWL fornece um amplo conjunto de modelagem que permite a um desenvolvedor de ontologias especificar mais ricamente as classes em termos das condições necessárias e suficientes para a adesão a essas classes (GIBBINS; SHADBOLT, 2009; SILVA et al., 2009; PASTOR SÁNCHEZ, 2012).

As ontologias oferecem uma grande capacidade de interoperabilidade semântica permitindo descrições complexas de objetos, assim como as relações lógicas entre os mesmos. Uma ontologia OWL permite organizar os recursos mediante uma taxonomia de classes, junto com propriedades de dados para representar atributos e propriedades de objeto para estabelecer as relações entre os recursos. Com OWL é possível definir os intervalos e domínios dessas propriedades e estabelecer restrições (PASTOR SÁNCHEZ; LLANES PADRÓN, 2017, p. 302, tradução livre).

Na prática, a OWL proporciona uma maior interpretação de conteúdos informacionais disponíveis na *Web* do que o suportado pelo XML, RDF e RDF *Schema*, uma vez que fornece um vocabulário adicional com uma semântica formal,

representando de forma mais ampla o significado dos termos e as relações existentes entre eles.

Jorente, Pádua e Santarém Segundo (2017, p. 9) destacam que, uma ontologia pode ser considerada como uma,

[...] moldura que visa normalizar as terminologias utilizadas para a estruturação dos recursos e agentes. Além da possibilidade de variar dentro das generalidades de certo termo, tornando possível a interoperabilidade e o acesso aos repositórios de dados (*Datasets*) de diferentes tipos que estão sendo apontados e acessados pelas linguagens e pelos recursos utilizados em dado projeto.

De acordo com o World Wide Web Consortium (2012), a segunda versão da recomendação da OWL não difere muito da original uma vez que, a função central de compatibilidade entre RDF/XML, o papel de outras regras de sintaxe e os relacionamentos semânticos não foram alterados. Com isso, cada ontologia serializada de acordo com essa especificação também pode ser serializada como um documento RDF em uma sintaxe concreta adequada, como RDF/XML, conforme ilustra a figura 11.

Figura 11 - Relações sintáticas da OWL

Ontologia OWL 2 escrita na serialização XML.

```

<Ontology ontologyIRI="http://example.org/carro.owl" ...>
  <Prefix name="owl" IRI="http://www.w3.org/2002/07/owl#" />
  <Declaration>
    <Class IRI="Carro" />
  </Declaration>
</Ontology>

```

Sintaxe RDF / XML

```

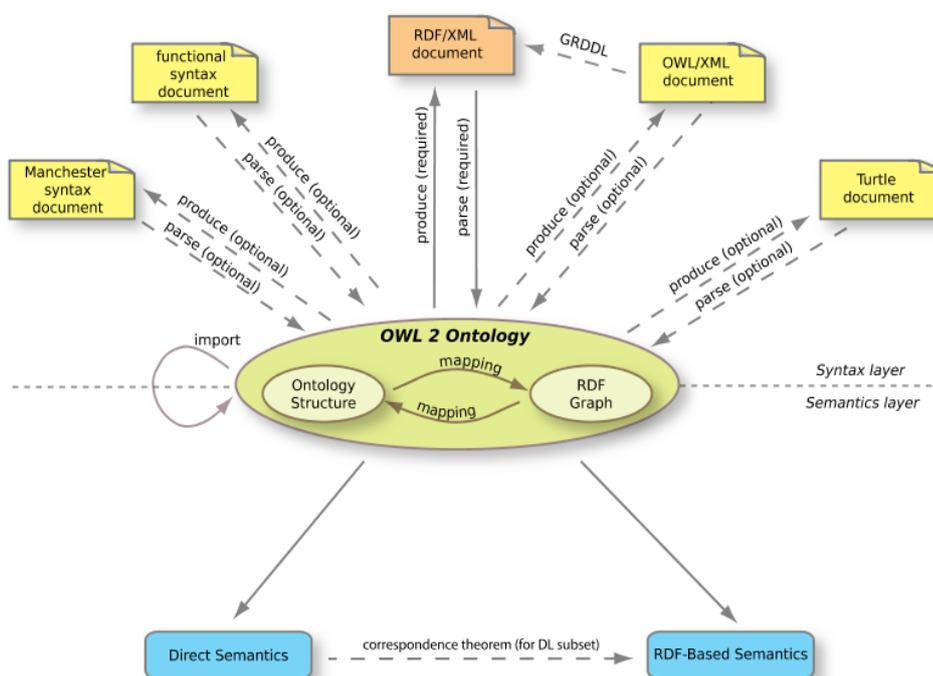
<rdf:RDF ...>
  <owl:Ontology rdf:about="" />
  <owl:Class rdf:about="#Carro" />
</rdf:RDF>

```

Fonte: Elaborado pelo autor

Na figura 12 é apresentada a estrutura da versão OWL 2, seus principais blocos de construção e como eles se relacionam entre si.

Figura 12 - Estrutura da OWL 2



Fonte: W3C (2012).

A elipse, no centro, representa a noção abstrata de uma ontologia, que pode ser pensada como uma estrutura abstrata ou como um grafo RDF. Na parte superior estão as várias sintaxes concretas (*Manchester*, *RDF/XML*, *OWL/XML*, *Turtle*) que podem ser usadas para serializar e intercambiar ontologias, sendo as mesmas opcionais ou obrigatórias. Já na parte inferior, têm-se duas especificações semânticas que definem o significado das ontologias da linguagem OWL 2.

Para o grupo de estudo do W3C, responsável pelo desenvolvimento da recomendação, a maioria dos usuários dessa linguagem necessitará somente de uma sintaxe, na parte superior, e uma semântica, na parte inferior, uma vez que qualquer ontologia OWL 2 pode ser visualizada como um grafo RDF.

Com isso, uma sintaxe concreta é necessária para armazenar e intercambiar uma ontologia OWL entre ferramentas e aplicativos, cuja sintaxe de intercambio primária é o *RDF/XML*, que é de fato a única sintaxe que deve ser suportada por todas as ferramentas dessa linguagem (WORLD WIDE WEB CONSORTIUM, 2012).

Freitas Jr. e Jacyntho (2016, p. 53) salientam que, a criação de uma ontologia tem por objetivo relacionar nomes a itens representados no qual a escolha da mesma “[...] deve ser adequada ao objetivo que se espera alcançar na questão de atribuir significados e o reconhecimento dos mesmos por sistemas que irão

interpretar as triplas RDF produzidas por uma aplicação”.

No quadro 2 é apresentado as principais especificações de sintaxes que são mandatórias e opcionais para a OWL 2.

Quadro 2 - Especificações das sintaxes da OWL

Nome da Sintaxe	Especificação	Status	Propósito
RDF/XML	Mapeamento para grafos RDF, RDF/XML	Obrigatório	Intercâmbio (pode ser escrito e lido por todos os softwares OWL 2 em conformidade)
OWL/XML	Serialização XML	Opcional	Mais fácil de processar usando ferramentas XML
Sintaxe Funcional	Especificação estrutural	Opcional	Mais fácil ver a estrutura formal das ontologias
Sintaxe Manchester	Sintaxe de Manchester	Opcional	Mais fácil de ler / escrever ontologias DL
Turtle	Mapeamento para grafos RDF, Turtle	Opcional, Não da OWL-WG	Mais fácil de ler / escrever triplas RDF

Fonte: Adaptado da W3C (2012).

Normativamente, a OWL 2 é definida por cinco documentos de especificação principais que descrevem sua estrutura conceitual, sintaxe de intercâmbio primário entre RDF/XML, duas semânticas alternativas baseadas em RDF e requisitos de conformidade. Ademais, três novas especificações foram adicionadas para descrever recursos opcionais que podem ser suportados por algumas implementações: os perfis de linguagem e duas sintaxes alternativas concretas (OWL/XML e *Manchester*²).

No quadro 3, apresenta-se os principais documentos e as especificações que podem ser utilizadas por implementadores, desenvolvedores e usuários da linguagem OWL 2.

² A sintaxe *Manchester* foi criada para lidar com os problemas lógicos e fornecer aos não-lógicos uma sintaxe que facilita a criação de ontologias. Foi projetado principalmente para apresentar e editar expressões de classe em ferramentas, mas também pode ser usado para representar ontologias completas. Maiores informações sobre a sintaxe de OWL Manchester, ver o artigo *The Manchester OWL Syntax*, disponível em: http://webont.org/owlled/2006/acceptedLong/submission_9.pdf

Quadro 3 - Especificações da OWL 2

Parte	Tipo	Documento
1	Para usuários	Visão geral – inclui uma descrição da especificação OWL 2 e sua relação com o OWL 1. Este é o ponto de partida e de referência para OWL 2.
2	Especificação principal	Especificação Estrutural e Sintaxe de Estilo Funcional – define as construções de ontologias OWL 2 em termos de sua estrutura e uma sintaxe de estilo funcional e, define ontologias de DL OWL 2 em termos de restrições globais.
3	Especificação principal	Mapeamento para grafos RDF – define um mapeamento das construções em grafos RDF e, assim, define os principais meios de intercâmbio de ontologias OWL 2 na Web Semântica.
4	Especificação principal	Semântica Direta – define o significado de ontologias OWL 2 em termos de uma semântica teórico-modelo.
5	Especificação principal	Semântica Baseada em RDF – define o significado das ontologias OWL 2 através de uma extensão da semântica RDF.
6	Especificação principal	Conformidade – fornece requisitos para ferramentas OWL 2 e um conjunto de casos de teste para ajudar a determinar a conformidade.
7	Especificação	Perfis – define três sub-linguagens da OWL 2 que oferecem vantagens importantes em cenários de aplicativos específicos.
8	Para usuários	OWL 2 Primer – fornece uma introdução acessível ao OWL 2, incluindo orientação para aqueles que vêm de outras disciplinas.
9	Para usuários	Novos Recursos e Justificativa – fornece uma visão geral dos principais (novos) recursos do OWL 2 e motiva sua inclusão no idioma.
10	Para usuários	Guia de referência rápida – fornece um guia simplificado para as construções da OWL 2, observando as mudanças da OWL 1.
11	Especificação	Serialização XML – define uma sintaxe XML para intercâmbio adequada para uso com ferramentas XML, como editores baseados em esquema e em XQuery/XPath.
12	Especificação	Sintaxe Manchester – define uma sintaxe fácil de ler, mas menos formal, usada em algumas ferramentas de interface de usuário, também disponível na versão Primer.
13	Especificação	Extensão do intervalo de dados: equações lineares – especifica uma extensão opcional para o OWL 2, que suporta restrições avançadas sobre os valores das propriedades.

Fonte: Adaptado da W3C (2012).

Na área de Ciência da Informação, quanto mais utilizadas e compartilhadas as ontologias no ambiente *Web*, maior será sua consistência e aperfeiçoamento, uma vez que podem ser reutilizadas em um dado domínio do conhecimento (RAMALHO, 2006; FREITAS JR.; JACYNTHO, 2016).

Com isso, dada as constantes atualizações da linguagem OWL, Ramalho (2006) destaca que as contribuições da área de Ciência da Informação podem ser significativas a partir de estudos interdisciplinares que permitam estruturar e organizar os recursos informacionais desenvolvidos para os ambientes digitais. No quadro 4 é apresentado as principais vantagens de uso das linguagens de representação, levando-se em consideração as especificidades de cada uma e suas relações com as práticas emergentes no âmbito da representação de informações na *Web*.

Quadro 4 - Atributos das Linguagens de Representação

SGML	HTML	XML	RDF	OWL
Eficaz para formatar todos os tipos de documentos	Exibe e imprime conforme apropriado para o dispositivo implementador	Realiza o intercâmbio de dados de forma simples	Maior capacidade de interoperabilidade de informações em ambientes digitais independente do domínio	Maior capacidade de interoperabilidade semântica
Eficaz para geração de todas as linguagens de marcação	Necessita de uma linguagem mais abstrata e semântica para tarefas complexas	Necessita de uma linguagem mais abstrata e semântica para tarefas complexas	Permite a interligação de dados representados em outros formatos	Define classes e propriedades e suas restrições
Muito complexo para o transporte de dados na internet	De fácil manuseio pelo usuário	Tem vantagem no transporte de dados na internet ao utilizar os protocolos HTTP	Explora a infraestrutura da <i>Web</i> a partir de URI/URL	Pode ser usado em conjunto com o RDF e XML
		O usuário pode definir o seu próprio conjunto de <i>tags</i>	Relaciona formalmente diversos tipos de documentos	Maior capacidade para descrever as propriedades dos recursos da <i>Web</i> e suas relações

		É habilitado para usar as aplicações anteriores	É a base para o processamento de metadados na Web no contexto do movimento <i>Linked Data</i>	Permite criar ontologias mais complexas
		Alguns arquivos XML requerem uma maior capacidade de memória e de processamento		Maior capacidade de interpretação de conteúdo através do uso de um vocabulário adicional e uma semântica formal

Fonte: Elaborado pelo autor.

O uso dessas linguagens para a criação de novas linguagens, modelos de representação, vocabulários e ontologias constituem ferramentas tecnológicas valiosas para a representação e recuperação de informações em um determinado domínio, pois com elas é possível aumentar a capacidade semântica dos dados representados na *Web* permitindo com que dados e informações sejam reutilizados de forma mais ampla nesse ambiente.

4 EVOLUÇÃO DOS FORMATOS BIBLIOGRÁFICOS DE REPRESENTAÇÃO

A noção de tecnologia pode ser compreendida como a “[...] percepção de um conjunto limitado de possíveis alternativas tecnológicas e do ideal de desenvolvimentos futuros” (DOSI, 1982, p. 152). Com isso, a criação de novas tecnologias de informação e a globalização da comunicação foram fatores primordiais para proporcionar à Ciência da Informação o desenvolvimento de novos modelos e métodos para a representação, transmissão, compartilhamento e recuperação de informações.

A estrutura do formato de comunicações foi aprovada como um modelo de Organização Nacional de Padrões de Informação (NISO) (Z39.2) e, depois, de Organização Internacional de Normalização (ISO) (2709), sendo adotada de maneira generalizada pelas bibliotecas.

Um sistema eficiente de recuperação de informação deve criar e manter um ou mais bancos de dados contendo registros pertencentes aos requisitos de determinada unidade informacional, cujo conteúdo da base de dados, os registros, devem conter informações relevantes para seus usuários como, por exemplo, as características de um determinado produto, componentes de uma aeronave, uma ferramenta, etc.

A criação e manutenção de um sistema de recuperação de informação requerem um conhecimento prévio do campo de atuação, dos usuários reais e potenciais e suas atividades, requisitos e interesses de informação. Outro fator a ser considerado em qualquer sistema de recuperação é a natureza bibliográfica do documento, que conterá detalhes de itens bibliográficos, com ou sem resumo e/ou texto completo e, informações pessoais e institucionais e assim por diante.

Dantas, Cordula e Araújo (2016, p. 43) destacam que, a partir do desenvolvimento de tecnologias usadas para a representação em ambientes digitais, criou-se uma fonte “inesgotável de informações” que possibilitou ao usuário desses sistemas a abertura da maior biblioteca pessoal e, tudo isso ao alcance de um simples click.

Ao longo dos últimos anos, várias tentativas foram feitas para desenvolver mecanismos para o intercâmbio de dados bibliográficos e, conseqüentemente, uma série de formatos de troca bibliográfica foram desenvolvidos. A maioria desses

formatos foi projetada para armazenar dados bibliográficos, embora alguns também possam acomodar outros tipos de dados. Dada à variedade de formatos bibliográficos disponíveis, pode ser difícil para o profissional de um sistema de recuperação de informação escolher o que melhor atenda às necessidades de uma determinada unidade informacional.

Para Ramalho, Zafalon e Ouchi (2014, p. 4133), a partir da criação de formatos bibliográficos usados para a transmissão e a transferência de informações em meios digitais “[...] um novo paradigma informacional é apresentado, com destaque ao acesso e não à propriedade, sendo possível ignorar barreiras geográficas, culturais e temporais”, no qual a rapidez e a eficácia na disseminação de documentos e conteúdos informacionais são algo imprescindível (ANDRADE; CERVANTES, 2012).

O principal objetivo dos formatos de metadados e das tecnologias de representação é permitir que as informações publicadas na *Web* sejam processadas e compreendidas por computadores de maneira significativa, proporcionando uma linguagem que relaciona os dados representados a objetos do mundo real, permitindo que uma pessoa ou um computador possa ter acesso às informações de maneira rápida e precisa (BERNERS-LEE, 2009; CUEVA YERBA; MEDINA BRAVO, 2015).

No entanto, para que a recuperação e uso de informações sejam efetivos os dados bibliográficos devem ser organizados adequadamente. A criação de registros bibliográficos de modo a facilitar a busca e a recuperação tanto local quanto por meio de ambientes digitais e, a troca de informações bibliográficas entre bibliotecas e/ou unidades de informação exigem formatos padronizados que sejam eficientes no processo de criação e troca de registros.

Zafalon (2008) destaca que, a existência de um padrão internacionalmente aceito minimiza a duplicidade de documentos e permite um compartilhamento mais amplo de registros bibliográficos entre sistemas.

Padrões de metadados que possuem uma estrutura semântica mais abrangente têm como principal objetivo possibilitar o intercâmbio de informações e dados em ambientes digitais de maneira rápida, segura e com a menor perda de dados possível.

No campo da Ciência da Informação, estudos são desenvolvidos com o intuito

de propor o aperfeiçoamento e/ou a criação de modelos e métodos mais eficazes para a representação de informações e dados, tornando a recuperação de conteúdo um processo satisfatório para os usuários de sistemas de informação.

Nessa perspectiva, destaca-se a seguir a evolução dos formatos bibliográficos que possibilitam maior capacidade de intercâmbio de dados ao permitir a vinculação de relacionamentos existentes entre os diversos recursos existentes no âmbito da *Web*.

4.1 Formato MARC

Desenvolvido na década de 1960, pela *Library of Congress* (LC), o formato MARC, acrônimo de *Machine-Readable Cataloging*, surge com a finalidade de automatizar os processos de catalogação e permitir a comunicação de descrições bibliográficas em um formato legível por computadores. Para Moreno e Brascher (2007) apesar de, à época, o formato estar sendo desenvolvido e pouco utilizado a padronização no armazenamento de informações e a recuperação desses conteúdos foram fatores contribuintes para o rápido desenvolvimento do formato.

Desde a sua criação, o formato MARC passou por atualizações importantes em sua estrutura, como a inclusão e/ou a modificação de seus campos e subcampos e, em 1969 foi lançado o MARC II, cuja estrutura apresentava a separação entre diretórios, subcampos e parágrafos.

Países como Canadá e França adaptaram o formato MARC e desenvolveram modelos para atender suas necessidades, criando assim diferenças entre os diversos formatos internacionais ao longo do tempo. A partir de então, com o intuito de estabelecer uma padronização entre formatos, promover a redução de custos e estabelecer uma maior consistência nas interpretações das regras locais de catalogação para facilitar o intercâmbio de informações bibliográficas, a *Library of Congress*, a *British Library* (UKMARC) e a *National Library of Canada* (CAN/MARC) decidem unir esforços obtendo como resultado o MARC 21, um formato que fornece um conjunto de códigos e designadores de conteúdo e diretrizes para gerenciar e formatar registros eletrônicos para recursos de informação.

O formato MARC 21 tem como objetivo codificar os dados gravados em cartões de catálogo em um formato legível por computador. Permite que dados de

catalogação disponíveis em uma biblioteca se comuniquem eletronicamente com outras instituições, no qual o formato de sua estrutura de comunicação possibilite o transporte de dados e a troca de informações entre sistemas, recursos esses que podem ser usados em uma variedade de ambientes de processamento.

Segundo a Library of Congress (2011), inicialmente ele era visto como um transportador para catalogar dados especificados pelo *Anglo American Cataloging Rules* (AACR) com um conjunto de componentes básicos: a estrutura do formato de comunicação, o conjunto de elementos de dados e os dados estruturados de acordo com os padrões de conteúdo. O MARC 21 é definido para os seguintes tipos de dados:

- **MARC 21 Format for Bibliographic Data**³: contém especificações de formato para codificação de elementos de dados necessários para descrever, recuperar e controlar várias formas de material bibliográfico, incluindo livros, folhetins, arquivos de computador, mapas, música, materiais visuais e materiais mistos;
- **MARC 21 Format for Holdings Data**: contém especificações de formato para codificação de elementos de dados pertinentes a acervos e informações de localização para todas as formas de material;
- **MARC 21 Format for Authority Data**: contém especificações de formato para codificação de elementos de dados relacionados a registros bibliográficos que podem estar sujeitos a controle de autoridade;
- **MARC 21 Format for Classification Data**: contém especificações de formato para codificação de elementos de dados relacionados a números de classificação e legendas associadas;
- **MARC 21 Format for Community Information**: fornece especificações de formato para registros contendo informações sobre eventos, programas, serviços e assim por diante, para que essas informações possam ser integradas no mesmo OPAC como dados em outros tipos de registro.

Zafalon (2012, p. 44) destaca que, por ser um padrão de metadados

³ Nesse trabalho será abordado somente o formato designado para Dados Bibliográficos.

largamente utilizado no domínio bibliográfico, principalmente em instituições internacionais, o uso do Formato MARC 21 Bibliográfico como “[...] padrão de estrutura de metadados descritivos tornam o registro, o gerenciamento e a recuperação dos bibliográficos mais eficientes”. Para a Library of Congress (2006), o Formato MARC 21 para Dados Bibliográficos define os códigos e convenções (*tags*, indicadores, subcampos e valores codificados) que identificam os elementos de dados em registros bibliográficos MARC.

Ao longo dos anos algumas implementações foram feitas no Formato MARC 21 para incorporar em sua estrutura um novo código de catalogação que atendesse o crescente desenvolvimento tecnológico e, dada à necessidade de migração de informações entre instituições que adotam diversos formatos foi desenvolvida uma versão do MARC 21 na estrutura XML, o que possibilita maior facilidade de manipulação do formato e permite a migração de dados entre sistemas com maior eficiência.

Desenvolvido em 2002, pela *Library of Congress* (LC) em conjunto com o *MARC Standards Office*, o MARC 21 XML Schema, ou simplesmente MARCXML, é uma estrutura flexível e extensível que permite trabalhar com dados MARC 21 de forma específica atendendo a necessidades distintas. Segundo Siqueira (2003), a diferença entre o formato MARC 21 e a sua versão em XML está na estrutura usada para organizar os dados bibliográficos e catalográficos.

Com isso, o MARCXML:

[...] é proposto como uma alternativa ao MARC 21 para codificar e transferir registros, utilizando a mesma organização lógica e significado dos campos que o compõe, de forma que não são necessárias alterações nos nomes e tags de campos, subcampos e indicadores (GARZÓN FARINÓS, 2014, p. 319, tradução livre).

O MARCXML é um esquema XML simples que contém informações MARC e trabalha com todos os dados codificados na forma original, permitindo que registros de dados passem por transformações adicionais (FLAMINO, 2006; CASTRO, 2008).

A principal proposta da LC com a criação do MARCXML era, “[...] solucionar a necessidade de flexibilidade, extensibilidade e modularidade sem, contudo, perder a especificidade de domínio, garantindo que a interoperabilidade entre sistemas não seja prejudicada” (ALVES, 2010, p. 74).

Com a possibilidade de extensão do formato e a flexibilidade estrutural como

características, o MARCXML permite uma interoperabilidade de dados mais eficiente uma vez que seus campos são compatíveis com o MARC 21, possibilitando a troca de informações de forma rápida com o mínimo de perda entre diversas unidades informacionais.

[...] a implantação efetiva do formato MARCXML como uma estrutura de representação de recursos informacionais da área da Ciência da Informação, ficariam mais eficientes considerando-se que o MARC, como um padrão de metadados completo e flexível, garante a qualidade do registro enquanto que a XML garante a interoperabilidade entre os sistemas de informação heterogêneos (FLAMINO; SANTOS, 2006, p. 12).

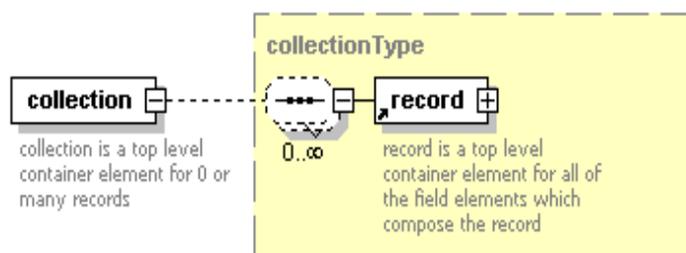
Para Siqueira (2003), a grande diferença entre o MARCXML e o seu antecessor, o MARC 21, é o fato da linguagem XML possuir uma estrutura que permite melhor organizar os dados bibliográficos e catalográficos, o que na prática representa uma melhor eficiência no intercâmbio e na recuperação de informações.

Na figura 13 é apresentado um fragmento do diagrama⁴ dos principais elementos MARCXML e o tipo de coleção a que pertence, como o <record>, <leader>, <controlfield>, <datafield> e o <subfield> e, sua representação no esquema.

⁴ Para maiores detalhes, consultar o diagrama completo do *Schema* MARC21slim.xsd, disponível no *Schema location*: <http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd> e *targetNamespace*: <http://www.loc.gov/MARC21/slim>

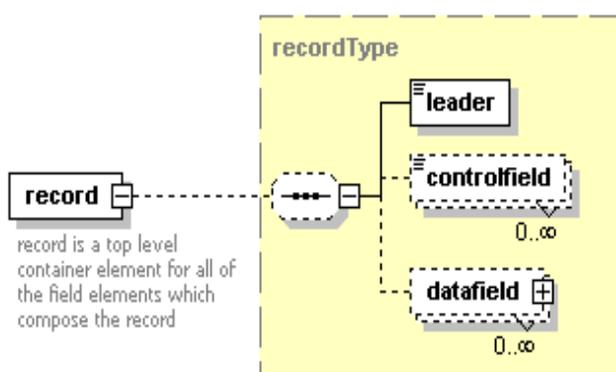
Figura 13 - Esquema MARCXML simplificado

Elemento Coleção e o tipo da coleção (Registro)



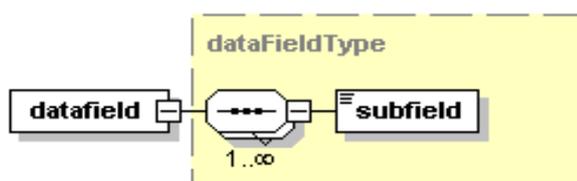
```
<xsd:element name="collection" type="collectionType" nillable="true" id="collection.e">
  <xsd:annotation>
    <xsd:documentation>collection is a top level container element for 0 or many records</xsd:documentation>
  </xsd:annotation>
</xsd:element>
```

Elemento Registro e seus tipos (Líder; Campo de Controle e Campo de Dados)



```
<xsd:element name="record" type="recordType" nillable="true" id="record.e">
  <xsd:annotation>
    <xsd:documentation>record is a top level container element for all of the field elements which compose the record</xsd:documentation>
  </xsd:annotation>
</xsd:element>
```

Elemento Campo de Dados e seu tipo (subcampo)



```
<xsd:element name="datafield" type="dataFieldType" minOccurs="0" maxOccurs="unbounded"/>
```

Fonte: Adaptado da Library of Congress, disponível em:
<http://www.loc.gov/standards/marcxml/xml/spy/spy.html>. Acesso em: 20 mar. 2017.

Considerado como um elemento de nível superior, o **<collection>** é o elemento raiz de um registro MARCXML e contém como elemento filho o **<record>**

que é usado para cada registro MARC que será incluído em um arquivo XML, podendo conter nenhum ou vários registros. O <record> é composto pelos seguintes elementos: <leader>, que contém o cabeçalho de um registro MARC; <controlfield>, elementos não vazios na qual cada elemento possui uma *tag* de atributo que contém a etiqueta ou código de três dígitos e, o <datafield>, elemento utilizado para os demais campos. Nesse elemento, os códigos dos subcampos são descritos como valor e código do atributo entre as *tags* de início e fim do elemento <subfield> (GARZÓN FARINÓS, 2014, tradução livre).

Para facilitar a conversão entre os formatos, a LC desenvolveu o kit de ferramentas MARCXML, que é um conjunto de programas Java que permite aos usuários efetuar a conversão do arquivo em formato MARC (incluindo conversão de conjunto de caracteres completo) e, outros formatos disponíveis na arquitetura MARCXML (LIBRARY OF CONGRESS, 2004a).

A validação dos dados em MARC 21 para o MARCXML se dá através de uma ferramenta de software externo ao esquema e o mesmo fornecerá três níveis possíveis de validação.

- validação básica de XML de acordo com o MARCXML *Schema*;
- validação da marcação MARC 21 (campo e subcampo);
- validação do conteúdo do registro MARC 21 (valores codificados, datas, horas, etc).

Ao utilizar a linguagem XML como estrutura para os registros MARC 21, o formato MARCXML permite criar de modo mais fácil ferramentas que serão usadas para consumir, manipular e converter dados MARC.

Segundo a Library of Congress (2004b, tradução livre), o MARCXML pode ser usado potencialmente para:

- representar um registro MARC completo em XML;
- servir como uma extensão de esquema para o METS (*Metadata Encoding and Transmission Standard*);
- representar metadados coletados nos processos de *Open Archive Initiative*⁵ (OAI);

⁵ A temática Iniciativa dos Arquivos Abertos é abordada com maiores detalhes por Flaminio (2006) em: MARCXML: um padrão de descrição para recursos informacionais em *Open Archives*. Disponível em: <<http://hdl.handle.net/11449/93694>>. Acesso em: 24 jan. 2017.

- permitir a descrição original de recursos na sintaxe XML;
- incluir metadados em XML que podem ser agrupados com um recurso eletrônico.

Quanto a sua aplicabilidade, a principal vantagem no uso de MARCXML dá-se pelo fato do esquema suportar todos os dados codificados do MARC, independentemente do formato, e possuir uma estrutura extensível que permite aos usuários desse padrão conectar e reproduzir partes de software para criar soluções personalizadas que atendam às suas necessidades.

A criação do formato MARC foi a parte mais importante da infraestrutura desenvolvida em 1970, sendo o mesmo adotado de forma generalizada e mundial. Segundo a Library of Congress (2011, tradução livre), inúmeras ferramentas e sistemas de código aberto foram construídos em torno do MARC 21 para permitir pesquisa, descoberta e exibição aprimorada de registros MARC, resultando em milhares de sistemas altamente desenvolvidos que trabalham com sistemas integrados de bibliotecas, redes e serviços auxiliares.

As ferramentas MARC e o suporte de sistemas desenvolvidos ao longo dos anos são altamente evoluídos e muito utilizados para reduzir os custos operacionais dos processos ao qual o formato se destina.

Na área de Ciência da Informação, o formato MARC permitiu o desenvolvimento de um ambiente de compartilhamento de registros altamente bem-sucedido e seu uso resultou em grandes economias de custo para as bibliotecas. No contexto atual, o MARC ainda continua sendo um formato largamente utilizado desempenhando um papel fundamental no amplo compartilhamento de registros de catálogo das bibliotecas entre diversas instituições.

No entanto, por ter uma estrutura complexa e extensa para os padrões da Internet, que dispõe de um grande volume de informações e que tende a aumentar cada vez mais, foi criado o padrão *Dublin Core*, considerado um padrão flexível e com o mínimo necessário de atributos em sua estrutura que surge com o objetivo de tornar a representação de recursos na *Web* um processo mais fácil e ágil.

4.2 Dublin Core

O padrão de metadados *Dublin Core* surgiu em 1995 a partir da chamada revolução da *Web* e das discussões em torno da *Web* semântica, com a proposta de ser um padrão simples auxiliando nos processos de acesso e recuperação de conteúdos informacionais disponibilizados na *Web*.

O DC foi desenvolvido a partir de *Workshops* realizados pela comunidade do *Dublin Core Metadata Initiative*⁶ (DCMI), que tinham como principal missão criar mecanismos que facilitassem “[...] a recuperação de recursos na Internet, utilizando-se de padrões de metadados” (GRÁCIO, 2002, p. 42).

Inicialmente, os primeiros *Workshops* em torno da discussão sobre o *Dublin Core* pretendiam popularizar a ideia de “metadados essenciais” para descrições de recursos simples e genéricos no âmbito da *Web*. Composto apenas por quinze elementos em sua estrutura, o padrão DC logo obteve ampla divulgação como parte do *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) – (Protocolo da Iniciativa dos Arquivos Abertos para a Coleta de Metadados), sendo ratificado como IETF RFC 5013, Norma ANSI/NISO Z39.85-2007 (atualmente na versão 2012) e Padrão ISO 15836:2009.

Os 15 elementos do formato *Dublin Core* fazem parte de um conjunto maior de vocabulários de metadados e especificações técnicas mantidas pelo DCMI. O conjunto completo de vocabulários, Termos de Metadados do DCMI (DCMI-TERMS), também inclui conjuntos de classes de recursos, além do vocabulário do DCMI-TYPE, esquemas de codificação de vocabulário e esquemas de codificação de sintaxe e, destinam-se para o uso em combinação com termos de outros vocabulários compatíveis no contexto de perfis de aplicação e com base no *DCMI Abstract Model* (DCAM) (DUBLIN CORE METADATA INITIATIVE, 2012, tradução livre).

A partir do ano 2000, a comunidade responsável pelos estudos relacionados ao DC concentrou esforços no desenvolvimento de perfis de aplicativos com o objetivo de que os registros de metadados passassem a utilizar o DC juntamente com outros vocabulários especializados para atender a implementação de requisitos

⁶ O *Dublin Core Metadata Initiative* (DCMI) é uma comunidade de estudiosos que além da criação do padrão *Dublin Core*, organizam regularmente *workshops* com a missão de promover a interoperabilidade e a harmonização de metadados, desde esquemas de elementos e vocabulários da *Web Semântica* até perfis de aplicativos, ontologias e sistemas de organização do conhecimento. Disponível em: <http://dublincore.org/>

específicos.

Desde então, com o constante desenvolvimento da WWW, o padrão DC tornou-se um dos vocabulários⁷ mais populares para uso com o RDF, favorecendo o uso de aplicações no contexto dados interligados.

A consolidação do modelo de dados RDF foi um dos fatores responsáveis que motivou esforços para traduzir o estilo de metadados de vocabulário misto da comunidade DC, em um modelo RDF compatível com o DCMI *Abstract Model*, modelo esse projetado para servir de ponte entre o paradigma moderno de gráficos de dados interligados ou não, com o paradigma mais familiar de registros de metadados validáveis, como os usados em *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) (POWELL et al., 2007).

Segundo Souza, Vendrusculo e Melo (2000), o *Dublin Core* é definido como um conjunto de elementos descritivos para metadados estruturados, cujo objetivo principal é o de facilitar a descrição de recursos eletrônicos, principalmente no ambiente *Web*.

O padrão de metadados Dublin Core é um conjunto de elementos simples e eficaz para descrever uma ampla gama de recursos em rede [...] cuja semântica foi estabelecida por consenso por um grupo internacional de profissionais interdisciplinar da biblioteconomia, da informática, da codificação de texto, da comunidade de museus e de outros campos de pesquisa relacionados (HILLMANN, 2001, não paginado, tradução livre).

Por vezes, o *Dublin Core* pode ser considerado como uma “[...] linguagem pequena para fazer uma classe particular de declarações sobre recursos” (BAKER, 2000, não paginado, tradução livre).

Segundo Baker (2000) e Hillmann (2001), o DC pode ser considerado uma linguagem que se aproxima, muitas vezes, das linguagens naturais, pois existem duas classes de termos, os elementos (substantivos) e os qualificadores (adjetivos), cuja sintaxe pode proporcionar a organização dos elementos e qualificadores em declarações em um padrão simples de afirmações.

Para Hillmann (2001, tradução livre), o DC possui algumas características que o torna um padrão internacionalmente aceito por diversas comunidades de usuários. São elas:

⁷ O termo vocabulários, utilizado aqui, refere-se ao conjunto de regras estabelecidas ao padrão DC para o uso com o modelo RDF.

- simplicidade de criação e manutenção – o conjunto de elementos do DC foi mantido tão pequeno e simples quanto possível para permitir que um não especialista crie registros descritivos simples para recursos de informação de forma fácil e barata, ao mesmo tempo em que fornece uma recuperação efetiva desses recursos no ambiente de rede.
- semântica comumente entendida – uma vez que a Internet é um vasto campo de informações com terminologias e práticas descritivas variadas, o DC pode auxiliar pesquisadores não especialistas a encontrar determinado item (artigos, obras, etc.) no universo de informações disponíveis na internet, uma vez que possui um conjunto comum de elementos cuja semântica é universalmente compreendida e apoiada.
- escopo internacional – originalmente desenvolvido no inglês, o conjunto de elementos do DC tem versões criadas em muitos outros idiomas, incluindo o finlandês, norueguês, tailandês, japonês, francês, português, alemão, grego, indonésio e espanhol. Outro objetivo do Grupo de Interesse Especial sobre o *Dublin Core* em Línguas Múltiplas, é ligar estas versões num registo distribuído utilizando a tecnologia RDF do *World Wide Web Consortium (W3C)*.
- extensibilidade – os elementos de metadados de outros conjuntos podem ser vinculados aos metadados do DC para atender à necessidade de extensibilidade. Este modelo permite que diferentes comunidades de usuários usem os elementos de DC para informações descritivas básicas que serão utilizáveis através da Internet, ao mesmo tempo em que permitem adições específicas de domínio que fazem sentido dentro de uma arena mais limitada.

Na figura 14 é apresentado um exemplo simples do DC que utiliza a linguagem RDF para expressar uma gravação de áudio. Em conjunto com a XML e o RDF, o DC pode ser utilizado com outros vocabulários de metadados e, nesse caso, usa o vocabulário controlado para descrever em detalhes a representação em questão.

Figura 14 - Representação do Dublin Core em RDF

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://media.example.com/audio/guide.ra">
    <dc:creator>Paulo Martins</dc:creator>
    <dc:title>Um guia sobre a catalogação</dc:title>
    <dc:description>Descreve os principais passos para o uso de vocabulários controlados.</dc:description>
    <dc:date>2017-03-27</dc:date>
  </rdf:Description>
</rdf:RDF>

```

Fonte: Elaborado pelo autor.

Os elementos que compõem a estrutura DC são subdivididos em elementos de conteúdo (*coverage*, *description*, *type*, *relation*, *source*, *subject* e *title*); propriedade intelectual (*contributor*, *creator*, *publisher* e *rights*), e de instanciação (*date*, *format*, *identifier* e *language*).

Em adição ao conjunto de elementos recomendados, o DCMI publicou em julho de 2000 a sua lista de Qualificadores, considerando duas grandes classes a seguir:

- **Element Refinement** – essas qualificadoras tornam o significado de um elemento mais específico. Um elemento refinado compartilha o significado do elemento não qualificado, mas com um escopo mais restrito. Por exemplo, um usuário que não entende um termo de refinamento de elemento específico deve ser capaz de ignorar o qualificador e tratar o valor de metadados como se fosse um elemento não qualificado (mais amplo);
- **Encoding Scheme** – essas qualificadoras identificam esquemas que auxiliam na interpretação de um valor de elemento. Esses esquemas incluem vocabulários controlados e notações formais ou regras de análise. Um valor expresso usando um esquema de codificação será assim um símbolo selecionado a partir de um vocabulário controlado (por exemplo, um termo de um sistema de classificação ou conjunto de cabeçalhos de assunto) ou uma sequência de caracteres formatada de acordo com uma notação formal (por exemplo, "2000-01-01" como a expressão padrão de uma data). Se um esquema de codificação não é compreendido por um usuário ou agente, o valor pode ainda ser útil

para um leitor humano (HILLMANN, 2001, não paginado, tradução livre).

O conjunto de elementos do DC pode ser representado em muitos outros formatos com a mesma especificação e sintaxe. A codificação mais utilizada permite que o DC codifique suas descrições como documentos HTML, XHTML, XML e RDF (BECKETT; MILLER; BRICKLEY, 2002; JOHNSTON; POWELL, 2008), no qual os documentos podem aparecer em RDF ou vinculados a outros documentos que os contenha (GARZÓN FARINÓS, 2014).

A codificação do padrão DC ocorre da seguinte forma:

- DC em XHTML – os metadados são incluídos no cabeçalho do documento XHTML, que utilizará a etiqueta <meta> no cabeçalho do elemento <head> onde serão incluídos os elementos DC a serem utilizados, <head profile="http://dublincore.org/documents/dc-html/">;
- DC em XML – na recomendação de 2008, o DCMI aborda os elementos simples e qualificados, na qual recomenda utilizar esquemas em vez de DTD. Os metadados DC devem aparecer em um elemento de campo com mesmo nível hierárquico, por exemplo:
<dc:subject>advogado</dc:subject>;
<dc:subject>escriturário</dc:subject>;
- DC em RDF/XML – baseado em um elemento de campo que utiliza o RDF, o vocabulário em XML é usado para o intercâmbio de metadados de qualquer tipo. No documento RDF se utilizará elementos próprios deste vocabulário, junto com os elementos do DC, cuja diferença está no nome de cada vocabulário. Nesse caso, o elemento raiz do documento RDF/XML <rdf:RDF> pode ter um ou mais elementos <rdf:description> para cada recurso descrito e, o elemento <rdf:description> irá conter o atributo “about” que terá um valor no *Uniform Resource Locator* (URL) ou no *Uniform Resource Identifier* (URI) que poderá conter os elementos DC ligados (GARZÓN FARINÓS, 2014, p. 344-347, tradução livre).

O padrão de metadados *Dublin Core* representa um vocabulário fundamental

para a *Web Semântica* e conseqüentemente para o *Linked Data* e é considerado um dos primeiros vocabulários a ser declarado como um conjunto de propriedades RDF e identificado pelo uso de URIs (BAKER, 2012, tradução livre), no qual o conjunto principal pode ser expandido com elementos adicionais sendo determinante para edição de páginas da *Web* possibilitando adição de *tags*, podendo ser modificado por qualquer usuário. Conforme ilustrado no exemplo da figura 15.

Figura 15 - Amostra de metadados do DC para uma página da web

```

<link rel="schema.DC" href="http://purl.org/dc">
<meta name="DC.Title" content="Paulo George Martins">
<meta name="DC.Creator" content="Paulo Martins">
<meta name="DC.Subject" content="Paulo George Martins; Departamento de
Ciência da Informação; Universidade Federal de Sao Carlos - UFSCar">
<meta name="DC.Description" content="Personal webpage">
<meta name="DC.Publisher" content="Universidade Federal de Sao Carlos - UFSCar">
<meta name="DC.Contributor" content="Ramalho">
<meta name="DC.Date" content="23/05/2018">
<meta name="DC.Type" scheme="DCMIType" content="Text">
<meta name="DC.Format" content="text/html 9865 bytes">
<meta name="DC.Identifier" content="http://www.ppgci.ufscar.br/pessoas/discentes">
<meta name="DC.Source" content="Personal files">
<meta name="DC.Language" content="Portugues">
<meta name="DC.Relation">

```

Fonte: Elaborado pelo autor

Considerado um padrão simples, flexível e extensível o DC permite uma melhor interoperabilidade semântica entre diferentes formatos e possibilita a recuperação de dados de forma eficaz, uma vez que pode ser codificado em linguagens que fundamentam a interligação de dados e informações em larga escala na Internet, como o RDF e a XML.

Tais características foram fundamentais para sua rápida aceitação e utilização no ambiente *Web*, tornando-o um padrão amplamente usado no contexto dos dados interligados.

4.3 *Bibliographic Framework* (BIBFRAME)

Lançado em 2011 pela *Library of Congress*, o *Bibliographic Framework* ou BIBFRAME, fornece uma base para a descrição bibliográfica contemporânea na *Web* ao ser fundamentado na proposta de dados interligados.

Segundo a Library of Congress (2016), o vocabulário BIBFRAME usa um modelo de dados interligados e potencializa a prática de modelagem RDF de identificação exclusiva como recursos da *Web* de todas as entidades, atributos e relacionamentos existentes entre essas entidades.

O BIBFRAME surge com a proposta de evoluir padrões de descrição bibliográfica para um modelo de dados semanticamente mais amplo, visando a reestruturação das descrições bibliográficas para o contexto dos dados interligados, tornando os recursos da biblioteca mais visíveis na *Web*, definindo uma nova maneira de representar e compartilhar dados bibliográficos (MCCALLUM, 2017).

É um modelo definido em RDF e usa várias ontologias de base para o RDF desenvolvido pelo W3C, como: OWL, RDFS e outros Sistemas de Organização de Conhecimento Simples e, foi desenvolvido dentro dos princípios e padrões exigidos pelos códigos e regras de catalogação das tecnologias emergentes de representação (ALVES; SANTOS, 2013).

Baseado nos princípios de dados interligados, o modelo BIBFRAME busca fornecer um padrão para modelar recursos bibliográficos tradicionalmente codificados no padrão MARC 21, utilizando-se da *Web* como uma arquitetura subjacente para moldar uma estrutura descritiva comum, permitindo uma integração muito maior dos recursos bibliográficos existentes e, criar um roteiro para avançar em direção ao refinamento e o desenvolvimento de abordagens alternativas (LIBRARY OF CONGRESS, 2012, tradução livre).

Embora seja um modelo que visa potencializar os processos de descrição bibliográfica e compartilhamento de recursos informacionais, das bibliotecas na *Web*, o BIBFRAME:

[...] deve ser capaz de acomodar qualquer número de modelos de conteúdo e implementações específicas, [...] permitir a troca de dados entre bibliotecas, [...] suportar novas regras de metadados e padrões de conteúdo [...] deve, portanto, ampliar e estreitar o universo do formato para a troca de dados bibliográficos (LIBRARY OF CONGRESS, 2012, p. 5, tradução livre).

Ramalho (2016, p. 303) destaca que, BIBFRAME é um modelo que oferece um “[...] impulso para o reposicionamento das bibliotecas no cenário contemporâneo, contribuindo para o fortalecimento de atividades relacionadas à identificação e formalização dos relacionamentos entre recursos disponíveis na *Web*”, permitindo que os processos de controle e troca de recursos informacionais sejam repensados

sob novas perspectivas.

O BIBFRAME está baseado na formalização dos relacionamentos existentes entre os recursos e não em registros isolados, deste modo possibilita explorar as relações de forma independente, evitando duplicidades de informações e favorecendo a interligação de recursos de naturezas distintas (RAMALHO, 2016, p. 299).

Segundo Arakaki et al., (2017, p. 4), ao adotar em sua estrutura a proposta de dados interligados para a descrição dos recursos, as “[...] entidades (obra, manifestação, registros de autoridades, como pessoas, entidades coletivas, assuntos) poderão ser descritas separadamente e uma única vez, bastando relacionar uma ou mais entidades”, tornando o compartilhamento de dados muito mais eficiente entre as unidades informacionais.

Projetado para integrar e promover a representação de entidades a partir do *framework* RDF, por meio do uso de URIs, o modelo BIBFRAME suporta informações mais ampla ao mesmo tempo que atende às necessidades muito específicas das bibliotecas, tendo como principais objetivos:

- diferenciar claramente o conteúdo conceitual e suas manifestações físicas (obras e instâncias);
- focar na identificação inequívoca das entidades de informação (autoridades);
- alavancar e expor relacionamentos entre entidades (LIBRARY OF CONGRESS, 2012, p.3, tradução livre).

Dada à escala mundial da *Web*, é imperativo poder citar dados de biblioteca de uma maneira que não apenas diferencie o trabalho conceitual (um título e autor) dos detalhes físicos sobre a manifestação da obra (números de página) mas, também identificar claramente as entidades envolvidas na criação de um recurso (autores, editores) e os conceitos (assuntos) associados a um recurso.

Nesse sentido, o modelo BIBFRAME tem como proposta incentivar a criação de entidades claramente identificadas e o uso de identificadores computacionais capazes de realizar interpretação dessas entidades.

A primeira versão do modelo BIBFRAME contém as seguintes Classes Principais: *Work* (Trabalho), *Instance* (Instância), *Authority* (Autoridade) e *Annotation* (Anotação). Qualquer uma dessas Classes pode ter subclasses, que podem ter suas

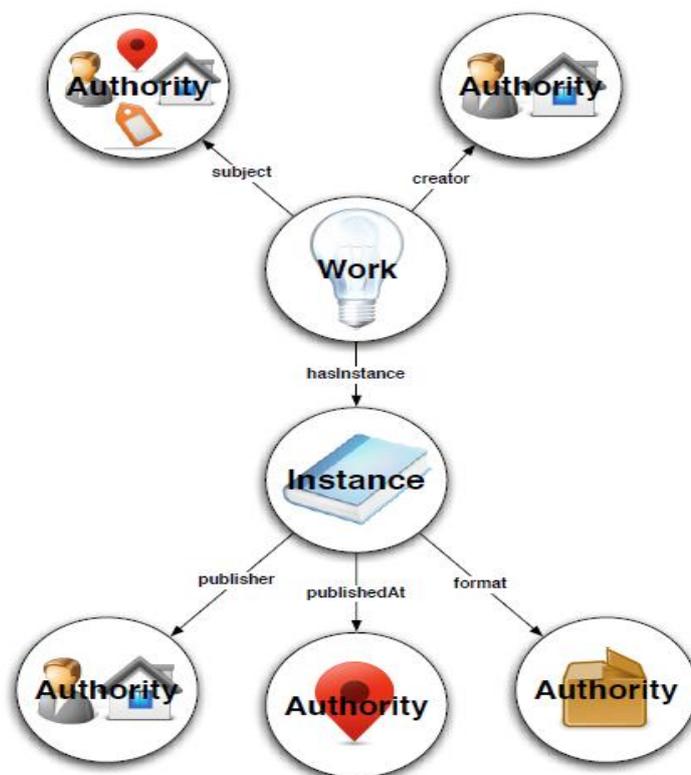
próprias propriedades.

- **BIBFRAME *Work*** – reflete uma entidade conceitual, abstrata, pois não há um objeto material único para o qual alguém possa representar. O *Work* existe como um ponto de controle baseado na *Web* que reflete a semelhança de conteúdo entre as várias instâncias associadas a Classe *Work*, bem como um ponto de referência. As propriedades comuns dessa Classe incluem relações contextuais com o BIBFRAME *Authorities* relacionadas à “sujeição” (tópico, pessoa, etc.) do trabalho, bem como as entidades (pessoa, organização, jurisdição, etc.) associadas à sua criação.
- **BIBFRAME *Instance*** – reflete uma personificação material e individual de um *Work* que pode ser de natureza física ou digital. Inclui propriedades específicas para as informações de materialização relacionadas à publicação, produção, fabricação e distribuição do material.
- **BIBFRAME *Annotation*** – afirma informações sobre um *Work*, *Instance*, *Authority* ou outra anotação. As anotações são declaradas para: expressar opiniões sobre um recurso, por exemplo, uma revisão; anexar informações específicas da instituição, por exemplo, participações; contribuir com aprimoramentos para uma descrição de recurso, por exemplo, capas de arte ou descrições resumidas.
- **BIBFRAME *Authority*** - é usado para identificar o que pode estar associado a um *Work* ou uma *Instance*, como: Agentes – pessoas, organizações, jurisdições, etc., com funções como autores, editores, distribuidores, etc.; Locais – áreas geográficas, como cidades, países, continentes etc.; Assuntos – conceitos atuais, conceitos temporais, lugares, agentes, etc. (LIBRARY OF CONGRESS, 2012, tradução livre).

Segundo a Library of Congress (2012), a Classe *Authority* não é projetada para substituir os esforços de autoridade existentes, mas sim fornecer uma camada de abstração leve e comum sobre vários esforços diferentes de autoridade baseada na *Web* para permitir o uso de dados de autoridade existentes. Conforme

apresentado na figura 16.

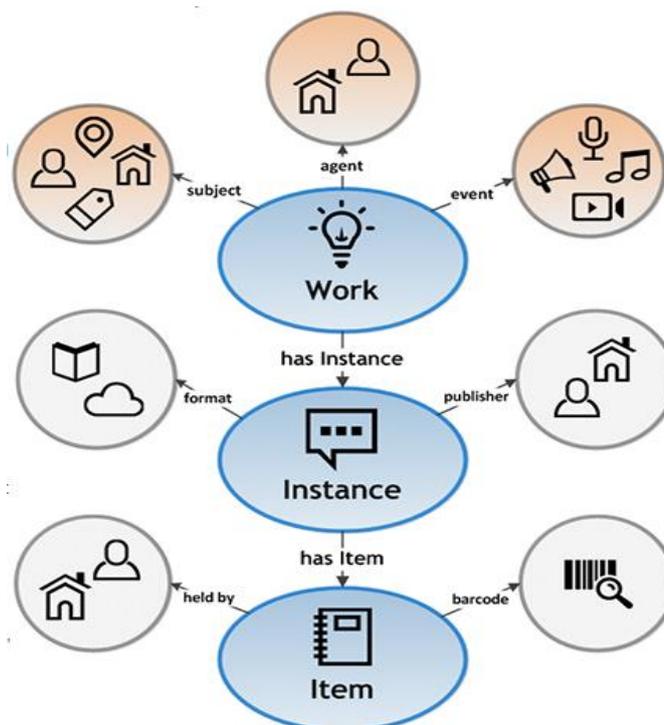
Figura 16 - Modelo BIBIFRAME 1.0



Fonte: Library of Congress (2012).

Em 2016, o modelo BIBFRAME passou por uma atualização em sua estrutura para refletir algumas melhorias necessárias ao projeto detectado na primeira versão, como por exemplo, a substituição das Classes *Authority* e *Annotation* pelas Classes *Item* e *Event*, sendo denominado de BIBFRAME 2.0, conforme apresentado na figura 17.

Figura 17 - Modelo BIBIFRAME 2.0



Fonte: Library of Congress (2016).

O vocabulário BIBFRAME consiste em classes e propriedades RDF, cujas classes são divididas em três classes principais de abstração, *Work* (Trabalho), *Instance* (Instância) e *Item*, e outras adicionais, muitas das quais são subclasses das principais. Já as propriedades descrevem características do recurso que está sendo descrito, bem como relacionamentos existentes entre eles (LIBRARY OF CONGRESS, 2016, tradução livre).

- **Work** – o mais alto nível de abstração e reflete a essência conceitual do recurso catalogado: autores, idiomas e assuntos;
- **Instance** – caracteriza as formas de realizações individuais de uma obra e reflete informações como, editor, local e data de publicação e formato;
- **Item** – é a cópia real (física ou eletrônica) de uma Instância e reflete informações como sua localização (física ou virtual), marca de prateleira e código de barras.

Além das três Classes principais, a Library of Congress (2012) destaca que o modelo ainda define conceitos adicionais que tem relação direta com as classes

principais, destacadas na figura. São eles: **Agents** (Agentes) considerado pessoas e organizações que são associados a um trabalho ou instância por meio de funções como autor, editor, etc.; **Subjects** (Assuntos), Trabalho (*Work*) que pode ser sobre um ou mais conceitos e inclui tópicos, lugares, eventos, itens, etc.; e, **Events** (Eventos), ocorrências, cuja gravação pode ser o conteúdo de um Trabalho (*Work*).

No modelo BIBFRAME 2.0 a classe de autoridade é eliminada, no entanto o componente controlado dos dados ainda contém as formas autorizadas de: (a) entidades nomeadas como pessoas, entidades coletivas, reuniões; (b) assuntos, incluindo tópicos, períodos de tempo e nomes geográficos; e (c) títulos uniformes cujos dados são incorporados na descrição do Trabalho BIBFRAME (XU; HESS; AKERMAN, 2017, p. 12, tradução livre).

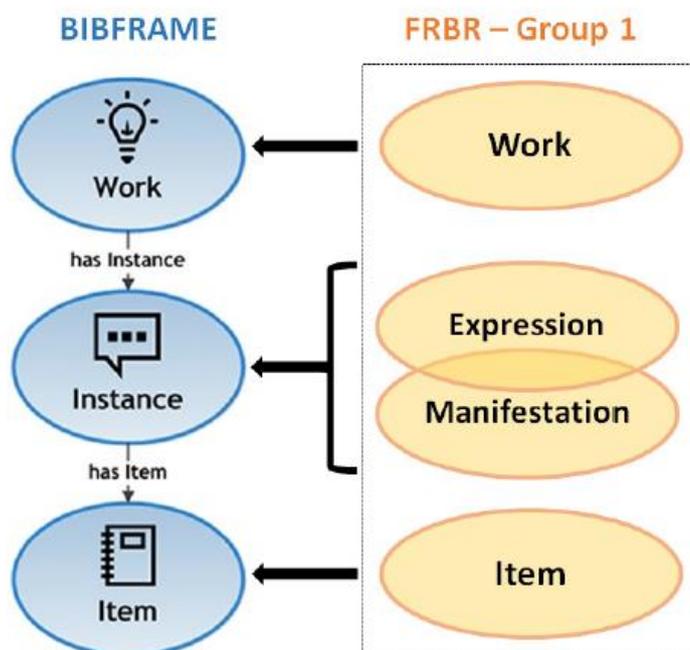
Segundo a Library of Congress (2012, tradução livre), o modelo de relacionamento de entidades, no qual inclui as regras preestabelecidas nos modelos conceituais *Functional Requirements for Bibliographic Records* (FRBR)⁸, ao utilizar as práticas de modelagem RDF o modelo BIBFRAME identifica de forma exclusiva como recursos da *Web* todas as entidades, atributos e os relacionamentos entre as entidades. Dessa forma, a classe *Work* se torna um ponto de controle na *Web* e reflete a semelhança de conteúdo entre as várias instâncias associadas à classe *Work*.

No modelo BIBFRAME, pode-se considerar a classe Instance como uma fusão das entidades Expression e Manifestation uma vez que, a origem dos aspectos das classes Work/Instance pode refletir os relacionamentos do FRBR, que se baseia nas as diretrizes do Resource Description and Access (RDA)⁹ em termos de um grafo RDF em vez de relacionamentos hierárquicos (SILVA et al., 2017), o que possibilita uma maior liberdade e simplificação das representações em ambientes digitais (RAMALHO, 2016; MCCALLUM, 2017). Conforme demonstrado na figura 18.

⁸ As diretrizes do *Functional Requirements for Bibliographic Records* não serão abordadas nesse trabalho, sendo citado apenas como referência às novas práticas adotadas no desenvolvimento de modelos de representação.

⁹ O código de catalogação *Resource Description and Access* (RDA), é um pacote de elementos de dados, diretrizes e instruções para a criação de metadados de recursos de bibliotecas e patrimônios culturais que são bem formados de acordo com modelos internacionais para aplicativos de dados interligados focados no usuário. Foi desenvolvido pelo Comitê Diretivo da RDA (anteriormente o Comitê Gestor para o Desenvolvimento da RDA) como parte de seu plano estratégico (2005-2009) para substituir as Regras de Catalogação Anglo-Americanas (AACR), 2ª Edição revisada, publicado pela primeira vez em 1978. Disponível em: <http://www.rda-rsc.org/content/about-rda>. Acesso em: 28 abr. 2018.

Figura 18 - Principais Classes do BIBFRAME e do Grupo 1 do FRBR



Fonte: Ramalho (2016).

Segundo Baker, Coyle e Petiya (2014); McCallum (2017), o BIBFRAME define e aproxima os dados bibliográficos da visão do FRBR/RDA, no qual a classe *Work* é aproximadamente equivalente ao FRBR *Expression* (Expressão) em combinação e, a classe *Instance* corresponde aproximadamente à *Manifestation* (Manifestação). A classe *Work* representa o conteúdo intelectual de um recurso e a classe *Instance* é a incorporação física de uma Obra e corresponde às manifestações FRBR/RDA.

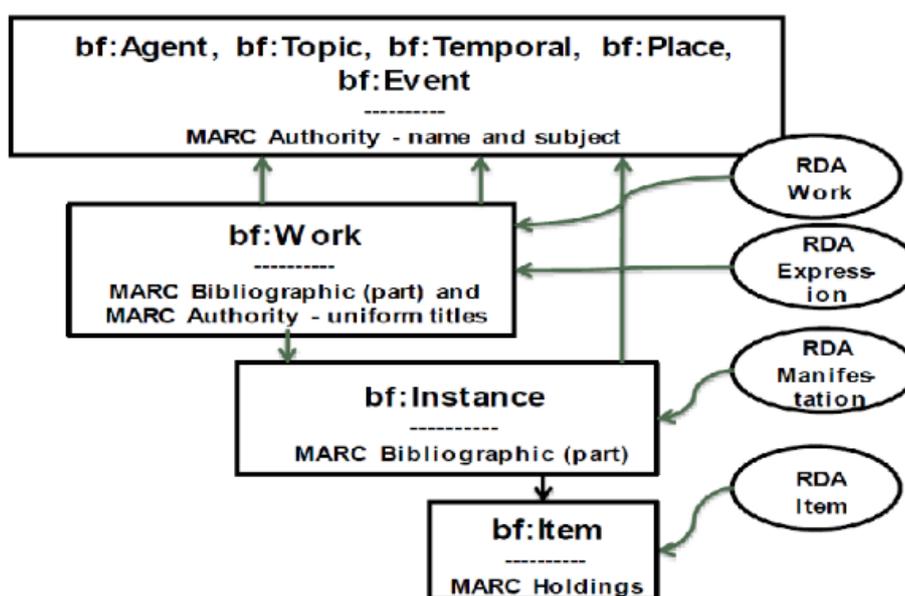
Para Arakaki et al., (2017. P. 2246), com o uso do modelo BIBFRAME “[...] a forma de catalogação proposta pelo FRBR é mais visível, pois, a catalogação poderá ser fragmentada e separada em blocos, ao invés de um registro único como era proposto pela estrutura do MARC 21”.

Tal iniciativa visa possibilitar maior capacidade de intercâmbio dos dados bibliográficos vinculados na *Web*, o que não é possível operacionalizar com o Formato MARC 21, uma vez que o mesmo foi projetado para “[...] atender às necessidades descritivas de catálogos manuais, transferindo aos computadores as informações contidas em fichas catalográficas (SILVA et al., 2017, p. 22).

Na figura 19 é apresentado um extrato do modelo BIBFRAME com referências ao mapeamento do padrão MARC e RDA. Nesse caso, as classes

BIBFRAME *Agent, Topic, Temporal, Place e Event*, comumente chamadas de dados de autoridade de nome e assunto, fornecem informações para conceitos-chave com relacionamentos definidos para as classes *Work* e *Instances*. A classe *Item*, “bf: Item” representa as informações típicas sobre itens que são mantidos nos campos originais.

Figura 19 - Modelo de dados BIBFRAME com mapeamento MARC e RDA



Fonte: McCallum (2017).

McCallum (2017, tradução livre) destaca que, existem algumas diferenças de MARC no modelo BIBFRAME na qual a mais proeminente é a modelagem de registros uniformes de autoridade de título do MARC como descrições da classe *Work* BIBFRAME. Para a autora, considerar a classe *Work* como Obra é apropriado e permite que ela tenha assuntos associados a essa classe. Com isso, as diferentes manifestações de um Trabalho são descritas separadamente como Instâncias e depois vinculadas à informação de Trabalho.

Todas as manifestações, Instâncias BIBFRAME, devem ter uma Obra à qual se relacionar, assim como no ambiente MARC era comum tentar descrever em um registro a manifestação e todas as diferentes portadoras para ele. Com a aplicação do modelo BIBFRAME, a expectativa é que grandes diferenças em suportes, como impressos e eletrônicos, sejam instâncias separadas de um trabalho e as características de cada uma possam ser claramente registradas nas descrições da

Instância (MCCALLUM, 2017, p. 80, tradução livre).

No site “<https://www.loc.gov/bibframe/mtbf/>” a LC disponibiliza os mapeamentos através do MARC para as Especificações de Conversão do BIBFRAME 2.0, que são agrupadas por campos MARC. As Especificações de Conversão MARC para BIBFRAME 2.0 são escritas pela *Library Network Development* e o *MARC Standards Office*, de modo que cada elemento no MARC será pelo menos considerado, mesmo que não seja convertido.

Nas especificações de conversão de MARC para o modelo BIBFRAME, ocasionalmente é necessário criar algumas extensões ou fazer ajustes no Vocabulário BIBFRAME 2.0 para acomodar dados do MARC, disponível no site “<http://www.loc.gov/bibframe/docs/>”.

Os programas necessários para a conversão de dados MARC para o modelo BIBFRAME são escritos em XSLT e estão disponíveis para *download* no site do *Github* da *Library of Congress* em “<https://github.com/lcnetdev/marc2bibframe2>”.

A LC também compartilha componentes e ferramentas do BIBFRAME 2.0 com a comunidade bibliotecária, permitindo que as instituições interessadas no modelo possam converter seus arquivos bibliográficos para o BIBFRAME (Modelo e Vocabulário), facilitando a conversão para estruturas diferentes ou em combinação com o BIBFRAME.

Desde o lançamento do modelo BIBFRAME seus desenvolvedores mantêm atualizações frequentes para o vocabulário BIBFRAME e o código de transformação do MARC. Assim, a LC trabalha no desenvolvimento de um novo projeto utilizando o BIBFRAME cujo objetivo é criar um ambiente no qual um catalogador possa trabalhar sem referência ao sistema MARC. Tal projeto tem duas partes importantes: a primeira é a conversão de todos os arquivos bibliográficos da *Library of Congress* em MARC para BIBFRAME, na qual o catalogador criará descrições de novos recursos e, a outra é a criação de um recurso de descrição BIBFRAME que interage com o catálogo do modelo BIBFRAME (KROEGER, 2013; MCCALLUM, 2017).

Uma das etapas mais importantes para a transição de registros bibliográficos para o BIBFRAME e a interligação de dados, envolve o mapeamento de campos e subcampos MARC em modelos de dados e a reformatação de dados necessária para se adequarem as regras propostas no modelo (XU; HESS; AKERMAN, 2017, tradução livre).

Ao utilizar as linguagens RDF e XML em sua estrutura, o modelo BIBFRAME propõe ir além das limitações bem documentadas do formato de codificação MARC. Segundo Kroeger (2013, tradução livre), talvez uma das razões pela qual o formato MARC tenha suportado tanto tempo seja pelo motivo de estar alinhado com as regras de catalogação atuais e o desenvolvimento do RDA que, como complemento separado de qualquer formato de codificação pode ser implementado em um ambiente MARC promovendo a impressão de que o mesmo ainda funciona bem o suficiente para executar as funções exigidas pelas novas regras.

Na área da Ciência da Informação, as contribuições do modelo BIBFRAME dá-se na possibilidade de tornar os diversos recursos disponíveis nas bibliotecas mais visíveis na Internet dada a grande capacidade de intercâmbio de dados e informações que o mesmo proporciona em sua estrutura ao utilizar a linguagem RDF para potencializar e operacionalizar a interligação de recursos nesse ambiente. Com isso, uma grande variedade de conteúdo poderia ser compartilhada de forma rápida a partir da adoção de um único modelo, o que reduziria o custo de seus processos uma vez que o BIBFRAME permite a especificação de múltiplos indicadores para uma entidade.

A estrutura na qual o BIBFRAME está fundamentado é compatível e interoperável com as linguagens de representação bibliográfica e, que tendem a tornar esse processo mais automatizado uma vez que podem processar uma grande variedade de recursos informacionais através de inferências. Assim, um dos principais benefícios em usar o URI em vez da *string* de rótulo de nome para identificar e vincular entidades é que as informações sobre a entidade estão disponíveis em RDF no local em que o URI está endereçado na *Web* (MCCALLUM, 2017).

Considerada uma estrutura de metadados alinhada com as novas práticas emergentes propostas para o movimento de dados interligados baseada em RDF e codificada em XML, o BIBFRAME representa uma escolha óbvia para substituir a codificação MARC.

No entanto, apesar das inúmeras iniciativas promovidas pela LC para permitir um maior uso do BIBFRAME seja por instituições parceiras, seja por estudiosos e desenvolvedores do projeto, ele ainda é um modelo em desenvolvimento e, a transição de registros bibliográficos em grande escala de um formato mundialmente

utilizado como o Formato MARC, requer estudos mais aprofundados que garantam sua eficiência, eficácia e segurança para as atividades desempenhadas nas unidades informacionais em um cenário com ampla diversidade de modelos de representação de recursos bibliográficos.

5 CONSIDERAÇÕES FINAIS

O grande volume informações disponíveis em ambientes digitais tem acarretado um aumento expressivo por pesquisas e a busca cada vez maior por conhecimento no qual a área de Ciência da Informação propõe, através da criação de novos métodos e modelos de representação o uso das tecnologias semânticas com o intuito de favorecer um melhor compartilhamento e recuperação de informações.

Com o avanço crescente das aplicações tecnológicas voltadas para a representação e recuperação de informações, os desafios para Ciência da Informação tendem a aumentar uma vez que, com o efeito da globalização do conhecimento a demanda por informação é cada vez maior e os recursos disponíveis para busca nem sempre conseguem atender a essa demanda informacional, dado a diversidade de ferramentas disponíveis.

Assim, em um contexto social cada vez mais caracterizado por interações virtuais os profissionais da informação se deparam com novos desafios na constante busca para atender as necessidades informacionais de uma nova geração de usuários que já nasceram imersos em ambientes computacionais.

A evolução das linguagens e formatos de metadados utilizados na representação da informação e o surgimento de novas tecnologias semânticas estabelece um novo cenário ao permitirem uma representação estruturada dos diversos tipos de conteúdo informacional produzidos para os ambientes digitais.

A estrutura desenvolvida para a representação de recursos existente nas linguagens de representação permite que aplicações com estruturas similares possam compartilhar dados e informações de qualquer tipo, sem que ocorra nesse processo prejuízos relativos a perdas de conteúdo durante o intercâmbio.

No que diz respeito ao crescente uso de novas tecnologias computacionais para o intercâmbio de dados e informações disponibilizados na *Web*, discutir metalinguagens, métodos e modelos de descrição em ambiente digital tem sido a forma de aproximar aspectos conceituais e técnicos do fluxo da informação nesses ambientes.

Compreender as implicações que envolvem tais tecnologias não se restringe somente ao seu uso, mas sim as relações teóricas e práticas que norteiam suas

aplicações no contexto das novas propostas de dados interligados a partir do estabelecimento de relações semânticas mais complexas que possam ser compreendidas por computadores da mesma forma que seriam feitas por humanos, aproximando tais conceitos ao das linguagens naturais.

Este estudo teve a proposta de apresentar um compêndio da evolução das linguagens de representação, contribuindo para uma maior discussão desta temática na área de Ciência da Informação, no intuito de favorecer uma melhor compreensão de tais tecnologias, seus reflexos e aplicações na área de Ciência da Informação, ratificando o caráter de complementação e análise mais robustas sobre o desenvolvimento dos processos de representação e recuperação de informação na área da CI.

Observa-se que tais tecnologias sempre tiveram relação direta com os instrumentos e práticas profissionais identificados no campo da CI, favorecendo melhorias nos processos de representação e estruturação de recursos informacionais.

No âmbito da representação de recursos informacionais, a linguagem XML tem destaque fundamental nesse processo ao permitir a padronização e a estruturação de documentos produzidos para ambientes digitais, estabelecendo regras para a criação de esquemas e padrões de metadados com maior capacidade semântica, além de permitir que novas linguagens sejam desenvolvidas com uma estrutura simples e interoperável viabilizando o compartilhamento de recursos em larga escala.

Constata-se que, a aplicação da linguagem XML nesse ambiente tem como principal objetivo descrever formalmente os elementos que compõem a estrutura interna de um recurso. Contudo, para formalizar as relações existentes entre recursos distintos é necessário a utilização do RDF que é uma linguagem que tem como foco principal a representação dos relacionamentos existentes entre recursos.

A partir do rápido crescimento da *Web* e o aumento de conteúdos nesse ambiente, o modelo RDF surge com a proposta de potencializar o intercâmbio de qualquer tipo de recurso permitindo formalizar as relações entre os diversos recursos em ambientes digitais.

Independente da forma e o domínio no qual o recurso está representado, o RDF é a estrutura principal que consolida o movimento de dados interligados,

favorecendo a recuperação e o compartilhamento de dados e informações na escala do fluxo informacional que o ambiente *Web* proporciona.

A flexibilidade e a versatilidade apresentadas nas linguagens RDF e XML, possibilitam oportunidades ampliadas para explorar diferentes maneiras de combinar metadados bibliográficos criando novas ferramentas com capacidade para conectar uma maior gama de recursos, sendo as mesmas peças fundamentais para o desenvolvimento de formatos de representação emergentes alinhados com as propostas de dados interligados.

Destaca-se nesse cenário o modelo BIBFRAME, uma iniciativa da LC que favorece o desenvolvimento de padrões de descrição bibliográfica no contexto de dados interligados, com capacidade de difundir a informação bibliográfica de unidades informacionais e bibliotecas, para o amplo contexto da *Web*.

A proposta inicial do BIBFRAME é que, paulatinamente, o mesmo possa substituir o MARC 21, um formato internacionalmente consolidado e utilizado. Atualmente na versão BIBFRAME 2.0, o modelo definido em RDF conta com recursos de dados interligados que envolvem o mapeamento de campos e subcampos de MARC, em modelos de dados como a reformatação de dados necessária para se adequarem as novas regras propostas pelos códigos de catalogação.

Observa-se que, apesar das várias iniciativas da LC em popularizar o BIBFRAME, o mesmo ainda está em fase de implementação sendo objeto de estudo de várias instituições internacionais e universidades, interessadas em sua aplicabilidade dada a possibilidade de migração de recursos bibliográficos do MARC 21 para um modelo semanticamente mais eficaz no âmbito da *Web*.

Nessa perspectiva, para que a utilização do BIBFRAME possa acontecer em larga escala, estudos mais aprofundados sobre o modelo, sobretudo em âmbito nacional, são necessários com o objetivo de corroborar a eficiência e segurança do modelo para atender as necessidades das unidades informacionais que ainda utilizam o MARC 21 em seus processos de representação.

Apesar da ampla abordagem, nos últimos anos, sobre as linguagens e formatos utilizados na representação e recuperação de informações e, suas aplicações práticas no contexto dos ambientes digitais, tecnologias emergentes como o RDF e o BIBFRAME e, suas relações com o movimento de dados

interligados ainda são timidamente exploradas em âmbito nacional.

Após a publicação do BIBFRAME 2.0, novas pesquisas têm sido desenvolvidas com o intuito de difundir as vantagens de sua aplicação no processo de migração de dados bibliográficos do formato MARC 21, sobretudo na comunidade científica internacional parceira da *Library of Congress*, uma vez que é um modelo que visa proporcionar maior semântica no processo de representação e descrição bibliográfica e, permite um compartilhamento estruturado de dados e informações entre sistemas operacionais.

Nesse sentido, torna-se imperativo a necessidade de tecnologias que possibilitem descrever formalmente os relacionamentos existentes entre recursos informacionais para o desenvolvimento de modelos utilizados no âmbito de dados interligados, como por exemplo: o emergente BIBFRAME.

Evidencia-se a necessidade de um melhor entendimento da evolução das tecnologias semânticas e dos conceitos que as fundamentam, para que pesquisas que abordem temas como o BIBFRAME, *Linked Data*; *Data Science*; Publicação Ampliada e *Web Semântica*, um modelo que visa proporcionar maior semântica no processo de representação e descrição bibliográfica, possam realmente ser fundamentadas nos princípios da área de Ciência da Informação, fornecendo subsídios teóricos para que tais tecnologias não sejam desenvolvidas a partir de um ‘vazio conceitual’, com o único objetivo de atender demandas de mercado, de modo que os avanços tecnológicos possam ser devidamente sedimentados no campo teórico.

5.1 Pesquisas Futuras

Constatou-se a necessidade de estudos que corroborem para um melhor entendimento sobre a relação entre o RDF e as novas tecnologias voltadas para a interligação de dados na *Web*, com o intuito de ratificar o caráter de complementação e análise mais robustas sobre o desenvolvimento dos modelos utilizados no processo de representação, recuperação e disseminação de conteúdos informacionais na área da Ciência da Informação, compreendendo os aspectos conceituais e práticos que permeiam o contexto da globalização do conhecimento e o efeito que tal processo tem sobre as novas demandas informacionais.

Ainda que o volume de conteúdos disponíveis na *Web* seja um ponto desafiador no que diz respeito à representação de informações para diversas áreas do conhecimento, é inegável o fato de que as novas tecnologias que buscam interligar os conteúdos semanticamente se constituem processos eficientes para a representação e o intercâmbio de informações em um contexto de alta demanda, o que tem contribuído para a recuperação dos mais variados recursos de forma rápida, sistêmica e padronizada em sistemas informacionais digitais.

O presente trabalho caracterizou-se como uma discussão da literatura com o objetivo de proporcionar subsídios teóricos para a área de Ciência da Informação, buscando identificar os principais conceitos e métodos inerentes às linguagens computacionais e suas aplicações para a representação, estabelecendo uma relação conceitual entre as práticas tradicionais e os novos modelos que visam à criação de estruturas que permitem a indexação simultânea de recursos através do uso de múltiplos vocabulários e a interligação de dados, favorecendo um ambiente totalmente interoperável entre sistemas computacionais.

Com isso, observa-se que pesquisas ainda são necessárias no sentido de propor um aprofundamento na temática acerca das tecnologias de representação e os modelos e formatos empregados para a representação e recuperação de informação e, suas aplicações demonstrando de modo prático a eficácia e relevância dessas tecnologias relacionadas ao movimento de dados interligados pautando-se em princípios técnicos, teóricos e metodológicos sedimentados na área de Ciência da Informação.

Por fim, ressalta-se a necessidade de adoção de metodologias que proporcionem uma melhor transição das práticas tradicionalmente utilizadas na Ciência da Informação para as novas tendências de publicação de dados interligados, permitindo o aprimoramento das habilidades dos profissionais da informação quanto à representação de dados e informações em ambientes digitais.

REFERENCIAS

- AGUIAR, F. L. de; KOBASHI, N. Y. Organização e representação do conhecimento: perspectivas de interlocução interdisciplinar entre ciência da informação e arquivologia. **Encontro Nacional de Pesquisa em Ciência da Informação**, v. 14, 2013. Disponível em: <<http://enancib.sites.ufsc.br/index.php/enancib2013/XIVenancib/paper/view/155>>. Acesso em: 28 ago. 2017.
- ALEIXO, D. V. B. S.; SANT'ANA, R. C. G.; RAMALHO, R. A. S. **Resource description framework (RDF): o impacto nas publicações na Ciência da Informação**. 2017. In: VII SECIN. Linkando informação, culturas e comunidades: reflexões teórico-práticas, 2017, Londrina. **Anais eletrônicos**. Londrina: UEL, 2017. Disponível em: <<http://www.uel.br/eventos/cinf/index.php/secin2017/secin2107/paper/viewFile/472/326>>. Acesso em: 08 nov. 2017.
- ALESSO, H. P.; SMITH, C. F. **Developing Semantic Web services**. Massachusetts: A K Peters, Ltd, 2005. 441 p.
- ALLEMANG, D.; HENDLER, J. **Semantic web for the working ontologist: modeling in RDF, RDFS and OWL**. United States: Morgan Kaufmann Publishers, 2008. 330 p.
- ALMEIDA, M. B. Uma introdução ao XML, sua utilização na internet e alguns conceitos complementares. **Ciência da Informação**, v. 31, n. 2, p. 5-13, maio/ago., 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n2/12903>>. Acesso em: 20 out. 2016.
- ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2010. Disponível em: <<http://hdl.handle.net/11449/103361>>. Acesso em: 25 jul. 2016.
- ALVES, R. C. V.; SANTOS, P. L. V. A. da C. **Metadados no domínio bibliográfico**. Rio de Janeiro: Intertexto, 2013. In: IX Encontro Internacional de Catalogadores e II Encontro Nacional de Catalogadores. Catalogação: do real ao virtual, 2013. **Anais eletrônicos**, 2013. Disponível em: <<http://www.abinia.org/catalogadores/52-194-1-PB.pdf>>. Acesso em: 20 jun. 2016.
- ANDRADE, M. C. de; CERVANTES, B. M. N. A contribuição da organização do conhecimento para a interoperabilidade semântica: alternativas para repositórios institucionais. **Informações @ Profissões**, v. 1, n. 1/2, p. 152-170, jul./dez., 2012. Disponível em: <<http://www.uel.br/revistas/uel/index.php/infoprof/article/view/14593/12261>>. Acesso em: 26 nov. 2016.
- ARAKAKI, F. A. **Linked Data: ligação de dados bibliográficos**. 2016. 146 f. Dissertação (mestrado) - Faculdade de Filosofia e Ciências, Universidade Estadual

Paulista Júlio de Mesquita Filho, Marília, 2016. Disponível em: <<http://hdl.handle.net/11449/147979>>. Acesso em: 12 mar. 2017.

ARAKAKI, F. A. et al. BIBFRAME: tendência para a representação bibliográfica na web. **Revista Brasileira de Biblioteconomia e Documentação**, v. 13, p. 2231-2249, dez. 2017. Disponível em: <<https://rbbd.febab.org.br/rbbd/article/view/995>>. Acesso em: 10 jan. 2018.

BAKER, T. A Grammar of Dublin Core. **D-Lib Magazine**, v. 6, n. 10, out., 2000. Disponível em: <<http://www.dlib.org/dlib/october00/baker/10baker.html>>. Acesso em: 10 jan. 2017.

BAKER, T. et al. Library Linked Data Incubator Group Final Report. **W3C Incubator Group Report**, out., 2011. Disponível em: <<http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025/>>. Acesso em: 20 nov. 2017.

BAKER, T. Libraries, languages of description, and linked data: A Dublin Core perspective. **Library Hi Tech**, v. 30, n. 1, p. 116-133, 2012. Disponível em: <<http://www.emeraldinsight.com/doi/abs/10.1108/07378831211213256/>>. Acesso em: 20 ago. 2015.

BAKER, T.; COYLE, K.; PETIYA, S. Multi-entity models of resource description in the semantic web: a comparison of FRBR, RDA and BIBFRAME. **Library Hi Tech**, v. 32, n. 4, p. 562-582, 2014. Disponível em: <<https://www.emeraldinsight.com/doi/abs/10.1108/LHT-08-2014-0081>>. Acesso em: 23 jan. 2017.

BARRETO, A. de A. A condição da informação. **São Paulo em Perspectiva**, v. 16, n. 3, p. 67-74, jul./set., 2002. Disponível em: <<http://repositorio.ibict.br/handle/123456789/173>>. Acesso em: 20 fev. 2016.

BAX, M. P. Introdução às linguagens de marcas. **Ciência da Informação**, v. 30, n. 1, p. 32-38, jan./abr., 2001. Disponível em: <<http://www.scielo.br/pdf/ci/v30n1/a05v30n1>>. Acesso em: 11 Set. 2016.

BECKETT, D.; MILLER, E.; BRICKLEY, D. Expressing Simple Dublin Core in RDF/XML. **DCMI Recommendation**, jul. 2002. Disponível em: <<http://www.dublincore.org/documents/dcmes-xml/>>. Acesso em: 14 fev. 2017.

BECKETT, D. et al. RDF 1.1 Turtle: terse RDF triple language. **W3C Recommendation**, fev. 2014. Disponível em: <<https://www.w3.org/TR/turtle/>>. Acesso em: 30 abr. 2018.

BERGMAN, M. **Linked Data**: connect distributed data across the web. 2017. Disponível em: <<http://linkeddata.org/>>. Acesso em: 12 jul. 2017.

BERNERS-LEE, T. **The web model**: information hiding and URI syntax. 1998. Disponível em: <<https://www.w3.org/DesignIssues/Model.html>>. Acesso em: 15 mar. 2017.

- BERNERS-LEE, T. **Linked Data**. 2009. Disponível em: <<https://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 06 mar. 2017.
- BISCHOF, S. et al. Mapping between RDF and XML with XSPARQL. **Journal on Data Semantics**, v.1, n. 3, p. 147-185, set. 2011. Disponível em: <<http://link.springer.com/article/10.1007/s13740-012-0008-7>>. Acesso em: 11 fev. 2017.
- BIZER, C. et al. **Linked data on the web**. 2008. Disponível em: <<https://dl.acm.org/citation.cfm?id=1367760>>. Acesso em: 10 out. 2017.
- BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data - the story so far. **International Journal on Semantic Web and Information Systems**, p. 1-22, 2009. Disponível em: <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>. Acesso em: 18 set. 2017.
- BOCCATO, V. R. C. Linguagem documentária na representação e recuperação da informação pela perspectiva sociocognitiva em ciência da informação. In: BOCCATO, V. R. C.; GRACIOSO, L de S. (Org.). **Estudos de linguagem em Ciência da Informação**. Campinas, SP: Ed. Alínea, 2011. p. 9-34.
- BRICKLEY, D.; GUHA, R. V. RDF vocabulary description language 1.0: RDF schema. **W3C Recommendation**, fev. 2004. Disponível em: <<https://www.w3.org/TR/2004/REC-rdf-schema-20040210/>>. Acesso em: 16 set. 2015.
- BRICKLEY, D.; GUHA, R.V. RDF Schema 1.1. **W3C Recommendation**, fev. 2014. Disponível em: <https://www.w3.org/TR/2014/REC-rdf-schema20140225/#ch_introduction>. Acesso em: 31 jan. 2017.
- BRÜGGEMANN-KLEIN, A.; WOOD. D. The Validation of SGML Content Models. **Mathematical and Computer Modelling**, vol. 25, n. 4, p, 73-84, fev. 1997. Disponível em: <[https://doi.org/10.1016/S0895-7177\(97\)00025-3](https://doi.org/10.1016/S0895-7177(97)00025-3)>. Acesso em: 08 nov. 2017.
- CAMPOS, L. F de B. Metadados Digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf.**, n.23, 2007. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/download/318/390>>. Acesso em: 18 fev. 2018.
- CAPURRO, R. Epistemologia e Ciência da informação. In: V Encontro nacional de pesquisa em ciência da informação, 5., 2003, Belo Horizonte. **Anais eletrônicos**. Belo Horizonte: Escola de Ciência da Informação da UFMG, 2003. Disponível em: <http://www.capurro.de/enancib_p.htm>. Acesso em: 05 maio 2017.
- CASTRO, F. F. de. **Padrões de representação e descrição de recursos informacionais 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 Júlio de Mesquita Filho, Marília, 2008. Disponível em: <<http://hdl.handle.net/11449/93689>>. Acesso em: 10 set. 2016.

CATARINO, M. E. Simple knowledge organization system: construindo sistemas de organização do conhecimento no contexto da web semântica. **Informação & Tecnologia**, v.1, n.1, jan./jun., 2014. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/itec/article/view/19307>> Acesso em 09 fev. 2017.

CUEVA YERBA, Z. B.; MEDINA BRAVO, S. K. Web semântica. **Repositorio Académico de la Universidad Peruana de Ciencias Aplicadas**. 2015. Disponível em: <<https://repositorioacademico.upc.edu.pe/handle/10757/622946>>. Acesso em: 10 set. 2017.

CYGANIAK, R.; WOOD, D.; LANTHALER, M. RDF 1.1: concepts and abstract syntax. **W3C Recommendation**, fev. 2014. <<https://www.w3.org/TR/rdf11-concepts/>>. Acesso em: 25 fev. 2017.

DANTAS, C. L. M.; CÓRDULA, F. R.; ARAÚJO, W. J. Análise da representação da informação em modelos entidade relacionamento com base em metadados. **Archeion Online**, v. 4, n. 1, 2016.

DAUM, B.; MERTEN, U. **Arquitetura de sistemas com XML**: conteúdo, processo e apresentação. Tradução de Daniel Vieira. Rio de Janeiro: Campus, 2002.

DESMARAI, N. ABCs of XML: the Librarian's guide to the eXtensible Markup Language. **Library Faculty and Staff papers**. 2000. Disponível em: <http://digitalcommons.providence.edu/facstaff_pubs/31>. Acesso em: 23 fev. 2017.

DOSI, G. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. **Research Policy**, v. 11, n. 3, p. 147-162, 1982. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.319.868&rep=rep1&type=pdf>>. Acesso em: 05 maio 2017.

DUBLIN CORE METADATA INITIATIVE. Dublin Core metadata element set. **DCMI Recommendation**, jun. 2012. Disponível em: <<http://www.dublincore.org/documents/dces/>>. Acesso em: 23 mar. 2017.

EDWARDS, M. **XML**: data the way you want it. [S. l.]: Microsoft, 1997.

GARZÓN FARINÓS, M. F. **El registro de autoridades personales tras la aparición del web**. 2014. 557 f. Tese (Doutorado) - Universitat Politècnica de València, Valência, 2014. Disponível em: <<http://hdl.handle.net/10251/46972>>. Acesso em: 12 abr. 2015.

FAULKNER, S. et al. HTML 5.2. **W3C Recommendation**, dez. 2017. Disponível em: <<https://www.w3.org/TR/html5/>>. Acesso em: 20 mar. 2018.

FERREIRA, J. A.; SANTOS, P. L. V. A. da C. O modelo de dados resource description framework (rdf) e o seu papel na descrição de recursos. **Informação & Sociedade**, v. 23, n. 2, p. 13-23, 2013.

FERREIRA, M. S. Linguagem e representação: considerações no universo da ciência informação. **Revista Digital de Biblioteconomia & Ciência da Informação**, v. 11, n. 3, 2013.

FLAMINO, A. N. **MARXML**: um padrão de descrição para recursos informacionais em Open Archives. 2006. 164 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2006. Disponível em: <<http://hdl.handle.net/11449/93694>>. Acesso em: 24 jan. 2017.

FLAMINO, A. N.; SANTOS, P. L. V. A. da C. MARXML para a OAI. In: **XIV Seminário Nacional de Bibliotecas Universitárias (SNBU)**, 2006, Salvador. **Anais Eletrônicos**. Salvador, 2006.

FREITAS JUNIOR, N.; JACYNTHO, M. D. de A. Um protótipo Linked data para catalogação semântica de publicações. 2016. **Perspectivas em Ciência da Informação**, v.21, n.4, p.48-65, out./dez. 2016. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2664>>. Acesso em: 14 ago. 2017.

FURGERI, S. O papel das linguagens de marcação para a Ciência da Informação. **TransInformação**, v. 18, n. 3, p. 225-239, set./dez. 2006. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/670>>. Acesso em: 18 mar. 2017.

GAO, J. **Linked Data based Enterprise Content Management**. 2011. 63 f. Dissertação (Mestrado em Web Technology) – Faculty of Engineering, Science and Mathematics, University of Southampton, United Kingdom, 2011.

GIBBINS, N.; SHADBOLT, N. **Resource Description Framework**. 2009. Disponível em: <https://www.academia.edu/2821151/Resource_Description_Framework_RDF_>. Acesso em: 28 jan. 2017.

GNOLI, C. **The ontological approach to knowledge organization**. In: II Seminário de Pesquisa Ontológica, 2009, Rio de Janeiro.

GRÁCIO, J. C. A. **Metadados para a descrição de recursos da Internet**: o padrão Dublin Core, aplicações e a questão da interoperabilidade. 2002. 127 f. Dissertação (Mestrado) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2002. Disponível em: <<http://hdl.handle.net/11449/93722>>. Acesso em: 01 mar. 2017.

GONDIM, L. M. P.; LIMA, J. C. **A pesquisa como artesanato intelectual**: considerações sobre método e bom senso. São Carlos: EdUFSCar, 2006.

HEITLINGER, P. **O guia prático da XML**: conceitos, exemplos, prática e aplicações da linguagem universal. Lisboa: Centro Atlântico, 2001. (Coleção Tecnologias).

HICKSON, I. et al. HTML5: a vocabulary and associated APIs for HTML and XHTML. **W3C Recommendation**, set. 2014. Disponível em: <<https://www.w3.org/TR/2014/PR-html5-20140916/single-page.html>>. Acesso em: 21 jan. 2016.

HILLMANN, D. Using dublin core. **DCMI Recommendation**, abr. 2001. Disponível em: <<http://www.dublincore.org/documents/2001/04/12/usageguide/#whatismetadata>>. Acesso em: 26 mar. 2017.

HJORLAND, B. The concept of subject in Information Science. **Journal of Documentation**, v. 48, n. 2, p. 172-200, jun. 1992.

HYLAND, B.; ATEMEZING, G.; VILLAZÓN-TERRAZAS, B. Best Practices for Publishing Linked Data. **W3C Working Group Note**, jan. 2014. Disponível em: <<https://www.w3.org/TR/ld-bp/>>. Acesso em: 02 abr. 2018.

JACOB, E. K. Ontologies and the semantic web. **Bulletin of the American Society for Information Science and Technology**, v. 29, n. 4, p. 19-22, abr./maio, 2003. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/bult.283>>. Acesso em: 18 fev. 2018.

JACOBS, I.; WALSH, N. Architecture of the World Wide Web. **W3C Recommendation**, dez. 2004. Disponível em: <<https://www.w3.org/TR/webarch/#URI-persistence>>. Acesso em: 07 mar.2017.

JAIN, P. **Linked open data alignment & querying**. 2012, 140 f. Tese (Doutorado) - Wright State University, Índia, 2012. Disponível em: <http://rave.ohiolink.edu/etdc/view?acc_num=wright1345575500>. Acesso em: 11 set. 2017.

JOHNSTON, P.; POWELL, A. Expressing Dublin Core metadata using HTML/XHTML meta and link elements. **DCMI Recommendation**, ago. 2008. Disponível em: <<http://www.dublincore.org/documents/dc-html/>>. Acesso em: 14 mar. 2017.

JORENTE, M. J. V.; PADUA, M. C.; SANTARÉM SEGUNDO, J. E. Criação de padrões na web semântica: perspectivas e desafios. **Em Questão**, v. 23, n. 3, p. 158-180, set./dez. 2017. Disponível em: <<http://seer.ufrgs.br/index.php/EmQuestao/article/view/70466>>. Acesso em: 10 fev. 2018.

KLYNE, G.; CARROLL, J. J. Resource Description Framework (RDF): concepts and abstract syntax. **W3C recommendation**, fev. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210>>. Acesso em: 15 set. 2015.

KROEGER, A. The Road to BIBFRAME: the evolution of the idea of bibliographic transition into a post-MARC future. **Cataloging & Classification Quarterly**, v. 51, n. 8, p. 873-890, set. 2013. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/01639374.2013.823584>>. Acesso em: 20 maio 2018.

KONTOPOULOS, E.; KRAVARI, K.; BASSILIADES, N. **Object-Oriented modeling of RDF schema ontologies**. 2007. Disponível em: <https://www.academia.edu/16254938/ObjectOriented_Modeling_of_RDF_Schema_Ontologies>. Acesso em: 24/01/2017.

LASSILA, O.; SWICK, R. R. Resource Description Framework (RDF): model and syntax specification. **W3C Recommendation**, fev. 1999. Disponível em: <<https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>>. Acesso em: 09 mar. 2017.

LE COADIC, Y. F. **A ciência da informação**. 2. ed. Brasília: Brique de Lemos, 2004.

LIBRARY OF CONGRESS. **MARCXML Design Considerations**. 2004a. Disponível em: <<http://www.loc.gov/standards/marcxml/marcxml-design.html>>. Acesso em: 16 mar. 2017.

LIBRARY OF CONGRESS. **MARCXML: uses and features**. 2004b. Disponível em: <<http://www.loc.gov/standards/marcxml/marcxml-overview.html>>. Acesso em: 16 mar. 2017.

LIBRARY OF CONGRESS. **MARC 21 Bibliographic**. 2006. Disponível em: <<http://www.loc.gov/marc/bibliographic/bdintro.html>> Acesso em: 06 jun. 2015.

LIBRARY OF CONGRESS. **A Bibliographic Framework for the Digital Age**. 2011. Disponível em: <<http://www.loc.gov/bibframe/news/framework-103111.html>>. Acesso em: 27 mar. 2018.

LIBRARY OF CONGRESS. **Bibliographic framework as a Web of data: Linked Data Model and supporting services**. 2012. Disponível em: <<http://www.loc.gov/bibframe/pdf/marcl-d-report-11-21-2012.pdf>> Acesso em: 10 abr. 2016.

LIBRARY OF CONGRESS. **Sustainability of Digital Formats Planning for Library of Congress Collections**. 2015. Disponível em: <<http://www.digitalpreservation.gov/formats/fdd/fdd000075.shtml>>. Acesso em: 26 ago. 2015.

LIBRARY OF CONGRESS. **Overview of the BIBFRAME 2.0 Model**. 2016. Disponível em: <<https://www.loc.gov/bibframe/docs/bibframe2-model.html>>. Acesso em: 26 fev. 2018.

LIBRARY OF CONGRESS. **MARC21 Schema**. Disponível em: <<http://www.loc.gov/standards/marcxml/xml/spy/spy.html>>. Acesso em: 20 mar. 2017.

MARCONDES, C. A. **HTML 4.0 fundamental**: a base da programação para a web. 2. ed. São Paulo: Érica, 2007.

MARCONDES, C. H. Linked data – dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na web. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 17, n. 34, p.171-192, maio/ago., 2012. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n34p171>>. Acesso em: 12 fev. 2018.

MARTINS, P. G. M.; RAMALHO, R. A. S. Evolução das linguagens de marcação: um estudo preliminar a partir do enfoque da área de Ciência da Informação. In: 3º Encontro Internacional Dados, Tecnologia e Informação. Perspectivas e interdisciplinaridades em Ciência da Informação, 2016, Marília. **Anais eletrônicos**. Marília: UNESP, 2016. Disponível em: <<https://goo.gl/Lec61v>>. Acesso em: 12 jan. 2017.

MARTINS, P. G. M.; RAMALHO, R. A. S. Tecnologias semânticas: fundamentos para o entendimento do conceito de Linked Data. 2017. In: VII SECIN. Linkando informação, culturas e comunidades: reflexões teórico-práticas, 2017, Londrina. **Anais eletrônicos**. Londrina: UEL, 2017. Disponível em: <<http://www.uel.br/eventos/cinf/index.php/secin2017/secin2107/schedConf/presentations>>. Acesso em: 12 jan. 2018.

MANOLA, F.; MILLER, E. RDF Primer. **W3C recommendation**. fev. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-primer-20040210>>. Acesso em: 16 set. 2015.

MCCALLUM, S. BIBFRAME Development. **JLIS.it**, v. 8, n. 3, p. 71-85, set. 2017. Disponível em: <<https://www.jlis.it/article/view/12415>>. Acesso em: 23 abr. 2018.

MEADOWS, A. J. **A comunicação científica**. Tradução Antônio Agenor Briquet De Lemos. Brasília: Briquet De Lemos, 1998.

MÉNDEZ, E. GREENBERG, J. Linked data for open vocabularies and HIVE's global framework. **El profesional de la Información**, v. 21, n. 3, maio/jun. 2012. Disponível em: <<http://recyt.fecyt.es/index.php/EPI/article/view/epi.2012.may.03/17916>>. Acesso em: 26 maio 2017.

MICROSOFT. **XML schema elements**. 2015. Disponível em: <[https://msdn.microsoft.com/en-us/library/ms256142\(v=vs.110\).aspx](https://msdn.microsoft.com/en-us/library/ms256142(v=vs.110).aspx)>. Acesso em: 14 ago. 2015.

MILLER, E. An Introduction to the Resource Description Framework. **D-Lib Magazine** (on-line). 1998. Disponível em: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Acesso em: 20 jan. 2017.

MILLER JUNIOR, R. **Teoria e problemas de XML**. Porto Alegre: Bookman, 2003.

MONTEIRO, L. L. P.; JACYNTHO, M. D. de A. Dados ligados. Digitalização. Documentos digitais. Metadados. RDF. Web semântica. **Transinformação**, v. 28, n. 2, p. 241-251, maio/ago. 2016. Disponível em: <<http://www.brapci.inf.br/index.php/article/view/0000020479/a3bf08192d0abc0fd05559c41b102655>>. Acesso em: 07 Ago. 2017.

MORENO, F. P.; BRASCHER, M. MARC, MARCXML e FRBR: relações encontradas na literatura. **Informação & Sociedade: Estudos**, v.17, n.3, p. 13-25, set./dez. 2007. Disponível em: <<http://www.brapci.inf.br/index.php/article/view/0000007615/772d60adea5b75a8090a027e6641e8a7>> Acesso em: 07 abr. 2015.

MOTA, D. A. R.; KOBASHI, N. Y. Web semântica e web pragmática: discussão crítica sobre versionamento na web e limites conceituais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 9, n. 2, 2016. Disponível em: <<http://www.brapci.inf.br/v/a/22987>>. Acesso em: 02 dez. 2016.

NEEDLEMAN, M. H. XML. **Serials Review**, vol. 25, n. 1, p. 117-121, 1999. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0098791399801427#!>>. Acesso em: 15 out. 2016.

NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação. **Informação & Informação**, v.1, n.2, p. 37-45, jul./dez. 1996. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1603>>. Acesso em: 12 ago. 2017.

OLIVEIRA, C. C. V. de. A interação de usuários com o catálogo on-line do Pergamum. **Revista Brasileira de Biblioteconomia e Documentação**, v.4, n.2, p. 73-88, jul./dez. 2009. Disponível em: <<https://rbbd.febab.org.br/rbbd/article/view/104/144>>. Acesso em: 08 jun. 2017.

OLIVEIRA, M. de. **Origens e evolução da Ciência da Informação**. In: Ciência da Informação e Biblioteconomia: novos conteúdos e espaços de atuação. Belo Horizonte: Editora UFMG, 2008.

OPARA, E.; SRIVASTAVA, M. Extensible Markup Language: an enterprise integration web delivery system. **Communications of the IIMA**, v. 3, n. 2, 2003. Disponível em: <<http://scholarworks.lib.csusb.edu/ciima/vol3/iss2/5>>. Acesso em: 12 fev. 2017.

ORTEGA, C. D. Ciência da Informação: do objetivo ao objeto. In: **El objeto de estudio de la Bibliotecología/Documentación/Ciencia de la Información: propuestas, discusión, análisis y elementos comunes**. México: UNAM, 2013, p. 151-171. (Coleção Epistemología de la Bibliotecología y Estudios de la Información).

PALMÉR, M. **Learning Applications based on Semantic Web Technologies**. 2012. 103 f. Tese (Doutorado) - School of Computer Science and Communication, Stockholm, 2012. Disponível em: <<http://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A564709&dswid=-7989>>. Acesso em: 09 ago. 2017.

PASTOR SÁNCHEZ, J. A. **Tecnologías de la web semántica**. Barcelona: Editorial UOC, 2012.

PASTOR SÁNCHEZ, J. A.; LLANES PADRÓN, D. Records in context: el camino de los archivos hacia la interoperabilidad semántica. **Anuario ThinkEPI**, v. 11, p. 297-304, jul. 2017. <Disponível em: <<http://eprints.rclis.org/31366/>>. Acesso em: 06 set. 2017.

PEMBERTON, S. XHTML2. **W3C Working Group Note**, abr. 2013. Disponível em: <<https://www.w3.org/MarkUp/#recommendations>>. Acesso em: 22 fev. 2017.

PEREIRA, E. C.; BUFREM, L. S. Princípios de organização e representação de conceitos em linguagens documentárias. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 10, n. 20, p. 21-37, nov., 2007. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2005v10n20p21/302>>. Acesso em: 18 jan. 2018.

PEREIRA, A. M.; RIBEIRO JR. D. I.; NEVES. G. L. C. Metadados para a descrição de recursos da internet: as novas tecnologias desenvolvidas para o padrão Dublin Core e sua utilização. **Revista ACB**, v.10, n.1, p. 241-249, jan./dez., 2005. Disponível em: <<https://revista.acbsc.org.br/racb/article/view/414>>. Acesso em: 25 fev. 2017.

POWELL, A. et al. DCMI abstract model. **DCMI Recommendation**, jun. 2007. Disponível em: <<http://www.dublincore.org/documents/abstract-model/>>. Acesso em: 23 mar. 2017.

QUIN, L. R. E. XML Schema. **W3C Recommendation**. 2015. Disponível em: <<https://www.w3.org/standards/xml/schema>>. Acesso em: 15 jan. 2017.

RAMALHO, R. A. S. **Web Semântica**: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação. 2006. 120 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2006. Disponível em: <<http://hdl.handle.net/11449/93709>>. Acesso em: 23 jan. 2014.

RAMALHO, R. A. S. Análise do Modelo de Dados SKOS: sistema de organização do conhecimento simples para a web. **Informação & Tecnologia**, v. 2, n. 1, p. 66-79, jan./jul., 2015. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/itec/article/view/25995/14680>>. Acesso em: 30 out. 2016.

RAMALHO, R. A. S. Bibframe: modelo de dados interligados para bibliotecas. **Informação & Informação**, v. 21, n. 2, p. 292-306, maio/ago., 2016.

Disponível em:

<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/26425>>. Acesso em: 12 fev. 2017.

RAMALHO, R. A. S.; VIDOTTI, S. A. B. G.; FUJITA, M. S. L. Web semântica: uma investigação sob o olhar da ciência da informação. **DataGramZero**, v. 8, n. 6, p. 1-18, dez., 2007. Disponível em:

<<http://www.brapci.inf.br/index.php/article/view/0000004791/f434755108f9f36100744a870906ab91/>>. Acesso em: 15 jun. 2016.

RAMALHO, R. A. S.; ZAFALON, Z. R.; OUCHI, M. T. Ontologia baseada nos FRBR: proposta de aplicação em catálogos online. In: XV ENANCIB - Além das nuvens: expandindo as fronteiras da ciência da informação, 2014, Belo Horizonte. **Anais eletrônicos**. Belo Horizonte: ECI/UFMG, 2014. Disponível em:

<<http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8/view>> Acesso em: 27 mar. 2015.

SANTOS, P. L. V. A. da C.; SANT'ANA, R. C. G. Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, v. 42, n. 2, p. 199-209, maio/ago., 2013.

Disponível em: <<http://revista.ibict.br/ciinf/article/view/1382/1560>>. Acesso em: 06 fev. 2017.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 41-62, jan./jun., 1996. Disponível em:

<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>>. Acesso em: 23 jan. 2015.

SCHREIBER, G.; RAIMOND, Y. RDF 1.1 Primer: **W3C Working Group Note**, jun.

2014. Disponível em: <<https://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/>>. Acesso em: 26 jan. 2017.

SENGUPTA, A.; DILLON, A. Extending SGML to accommodate database functions: a methodological overview. **Journal of the American Society for Information Science**, v. 48, n. 7, p. 629-637, jul., 1997. Disponível em:

<<https://repository.arizona.edu/handle/10150/105183>>. Acesso em: 17 fev. 2017.

SILVA, D. L. da. et al. Ontologias e unified modeling language: uma abordagem para representação de domínios de conhecimento. **DataGramZero**, v. 10, n. 5, out., 2009. Disponível em:

<<http://www.brapci.inf.br/index.php/article/view/0000008256/aa38ddb84d6b891d2f6cc49d6b363aaf>>. Acesso em: 10 set. 2017.

SILVA, L. C. da. et al. O código RDA e a iniciativa BIBFRAME: tendências da representação da informação no domínio bibliográfico. **Em Questão**, v. 23, n. 3, p. 130-156, set./dez., 2017. Disponível em:

<<http://seer.ufrgs.br/index.php/EmQuestao/article/view/69549>>. Acesso em: 10 set. 2017.

SILVA, R. E. da. **As tecnologias da Web Semântica no domínio bibliográfico**. 2013. 134 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2013. Disponível em: <<http://hdl.handle.net/11449/93653>>. Acesso em: 12 jan. 2018.

SIQUEIRA, M. A. **XML na Ciência da Informação: uma análise do MARC 21**. 2003. 134 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2003. Disponível em: <<http://hdl.handle.net/11449/93720>>. Acesso em: 12 fev. 2016.

SIMIONATO, A. C.; ARAKAKI, F. A.; SANTOS, P. L. V. A. da C. Descrição em bibliotecas, arquivos, museus e galerias de arte: linkando recursos e comunidades. **Informação & Informação**, v. 22, n. 2, p. 449-466, maio/ago., 2017. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/31464/22027>>. Acesso em: 08 jan. 2018.

SOUSA, B. P. de; FUJITA, M. S. L. Do catálogo impresso ao *on-line*: algumas considerações e desafios para o bibliotecário. **Revista ACB**, v.17, n.1, p. 59-75, jan./jun., 2012. Disponível em: <https://revistaacb.emnuvens.com.br/racb/article/view/822/pdf_71>. Acesso em: 08 nov. 2017.

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, vol. 29, n.1, p. 93-102, jan./abr., 2000. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652000000100010&lng=pt&nrm=iso>. Acesso em: 25 mar. 2017.

TAUBERER, J. **What is RDF and what is it good for?** 2014. Disponível em: <<https://github.com/JoshData/rdfabout/blob/gh-pages/intro-to-rdf.md>>. Acesso em: 26 set. 2017.

TEIXEIRA, E. C. **Utilizando XML para publicação de dados multimídia na Web**. 2002. 77 f. Dissertação (Mestrado em Computação) - Ciências Exatas e Tecnológicas, Universidade Federal de São Carlos, São Carlos, 2002. Disponível em: <<https://repositorio.ufscar.br/handle/ufscar/527>>. Acesso em: 15 jan. 2017.

TESCH JUNIOR, J. R. **XML Schema**. Florianópolis: Visual Books, 2002.

TREVIÑOS, A. N. S. **Introdução à pesquisa em ciências sociais: a pesquisa qualitativa em educação**. São Paulo: Atlas, 1987.

VICTORINO, M. C.; MEDEIROS, M. B. B.; SANTOS, R. F. dos. Arquitetura da informação como arcabouço para a integração da Tecnologia da Informação com os processos organizacionais e a informação. In: XII ENANCIB, 2011, Brasília. **Anais Eletrônicos**, Brasília, 2011. Disponível em: <<http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/2021/Arquitetura%20-%20Victorino.pdf?sequence=1>>. Acesso em: 23 nov. 2016.

WORLD WIDE WEB CONSORTION. OWL 2 Web Ontology Language. **W3C Recommendation**, dez. 2012. Disponível em: <<https://www.w3.org/TR/2012/REC-owl2-overview-20121211/#ref-owl-2-manchester-syntax>>. Acesso em: 20 out. 2017.

XU, A.; HESS, K.; AKERMAN, L. From MARC to BIBFRAME2.0: crosswalks. **Cataloging & Classification Quarterly**, v. 56, n. 2-3, nov., 2017. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/01639374.2017.1388326>>. Acesso em: 20 maio 2018.

ZSILAVECZ, G. **Rita+**: an SGML based document processing system. 1993. 93 f. Dissertação (Mestrado em Ciência da Computação) - Faculty of Science, University of Cape Town, Cape Town, 1993. Disponível em: <<http://hdl.handle.net/11427/13531>>. Acesso em: 17 fev. 2017.

ZAFALON, Z. **Formato MARC 21 bibliográfico estudo e aplicações para livros, folhetos, folhas impressas e manuscritos**. São Carlos: EdUFSCar, 2008.

ZAFALON, Z. R. **Scan for MARC**: princípios sintáticos e semânticos de registros bibliográficos aplicados à conversão de dados analógicos para o formato MARC 21 bibliográfico. 2012. 169 f. Tese (Doutorado) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2012. Disponível em: <<http://hdl.handle.net/11449/103386>>. Acesso em: 11 nov. 2015.