

DOI: 10.5748/9788599693124-13CONTECSI/PS-4142

IDENTIFYING BASIC CONCEPTS OF RESOURCES INHERENT IN THE SEMANTIC WEB

Marcos Luiz Mucheroni (Universidade de São Paulo, São Paulo, Brasil) - mmucheroni@usp.br

José Fernando Modesto da Silva (Universidade de São Paulo, São Paulo, Brasil) - fmodesto@usp.br

Francisco Carlos Paletta (Universidade de São Paulo, São Paulo, Brasil) - fcpaletta@usp.br

The World Wide Web, is basically a system of Internet servers that support specially formatted documents. The documents are formatted in a markup language called HTML (HyperText Markup Language) that supports links to other documents, as well as graphics, audio, and video files. The objective of this work is to describe the evolution and Web contributions, through the analysis of some of its key concepts that refer to the Organization, presentation, availability and location of content, in particular, the concept of metadata, highlighting descriptive metadata that are specific to information science. The text is characterized as descriptive to list and define the terms related to the Web.

Keywords: Web, Semantic Web, HTML, XML, Metadata

IDENTIFICANDO CONCEITOS BÁSICOS DOS RECURSOS INERENTES À WEB SEMÂNTICA

A World Wide Web, é basicamente um sistema de servidores de Internet que suportam documentos especialmente formatados. Os documentos são formatados em uma linguagem de marcação chamada HTML (Hypertext Markup Language) que suporta links para outros documentos, bem como gráficos, áudio e arquivos de vídeo. O objetivo do trabalho é descrever a evolução e contribuições da Web, por meio da análise de alguns dos seus principais conceitos que se referem à organização, apresentação, disponibilização e localização de conteúdo, em especial, o conceito de metadados, destacando os metadados descritivos que são próprios da Ciência da Informação. O texto caracteriza-se como descritivo ao elencar e definir os termos relacionados com a Web.

Palavras-Chave: Web; Web Semântica; HTML; XML; Metadados

1 INTRODUÇÃO

Desde sua origem, a Web se apresenta como uma publicação de conteúdo, inicialmente buscava dar visibilidade apenas a artigos técnicos e científicos, assim como, ter um espaço de documentação universal dos dados, embora isto não estivesse explicitamente descrito na concepção do ambiente e da significação da “marcação” adotada (Berners-Lee, 1986), uma vez que é uma “Linguagem de Marcação de Dados” (Hiperlink Markup Language), também conhecida como Linguagem de Marcação de Hipertexto – HTML (HyperText Markup Language).

A Web também é estruturalmente uma linguagem interpretadora que executa como um aplicativo, a troca de mensagens entre um servidor que contém as páginas originais, e um ou mais clientes que acessam estas páginas, criam um processo para cada cliente, estabelecendo um sincronismo estes e seus processos.

A “linguagem interpretada” é uma linguagem de programação, onde o código fonte contido nessa linguagem é executado por um programa de computador chamado “interpretador”, que em seguida é executado pelo sistema operacional ou processador (Scott, 2006), esta independência do sistema operacional a torna passível de ser executada em trechos ou determinados pedaços.

No caso da linguagem HTML, isto significa que cada trecho, por exemplo, `<cmd>` e `</cmd>` é interpretado e executado sendo o “cmd” um comando para determinado contexto. Embora, desde o início, o HTML já contemplasse metadados inseridos em seus códigos de comandos, a significação de contextos maiores (cada cmd pode ter vários comandos aninhados), com o crescimento dos usuários percebeu-se a necessidade de uma Web Semântica.

Os metadados incluídos no HTML são chamados atributos, por exemplo, em uma imagem temos `<IMG SRC= "oCristo.gif" ALT= "Cristo Redentor" >`, Alt é um atributo que aparece ao deslizar o cursor sobre a imagem cujo nome é “oCristo” no formato gif. Há diversos atributos contidos na linguagem para descrição de metadados.

Assim como o HTML, o XML (eXtended Markup Language) surgiu de uma linguagem de marcação padrão pioneira chamada SGML, mas que tivesse maior flexibilidade para permitir uma marcação semântica e conservando o estilo HTML.

Neste sentido, inicialmente surgiram a URI (Uniform Resource Identifier) que pode ser pensada como um identificador único na Web, e que contém como um localizador URL (Uniform Resource Locator) ou um nome URN (Uniform Resource Name), ou ainda como ambos, para identificação, localização e acesso aos conteúdos.

Embora o processo de identificação seja simples, é o principal qualificador da Web em sua fase denominada Web 1.0, ao permitir a identificação por hiperlinks. Muitas publicações técnicas, especialmente a que discutem os padrões produzidos pela Internet Engineering Task Force - IETF e pelo World Wide Web Consortium - W3C, não explicam com clareza o termo URL, e uma vez que não é dada a necessária distinção entre URL e URI, isto faz com que em contextos técnicos e de softwares, na Web Global, o termo URL torne-se em muitos casos ambíguos, por exemplo, quando há a duplicação de localização dos documentos na Web.

Com o crescimento da Web, Tim Berners-Lee, Ora Lassila e James Hendler publicam o reconhecido texto: Web Semantic (Berners-Lee; Lassila; Hendler, 2001) no qual fornecem pistas para a emergência de uma nova Web mais estruturada.

Em meados do ano de 2005, tendo havido a construção de muitas ferramentas, surge uma Web mais acessível à “publicação” de conteúdos pelos usuários, a denominada: Web 2.0.

A partir de então surge uma camada interna e pouco visível da Web, composta pelos marcadores de conteúdo XML (Extended Markup Language), e que também podem coexistir com os metadados destinados à organização dos dados, e fornecendo agilidade ao processamento destes dados. Possibilitando a “busca” por mecanismos buscadores, muitas vezes chamadas de inteligentes, mas que na prática precisaram de mais uma camada para que isto se tornasse realidade.

Para formatar o XML é criada uma gramática, o DTD (Data Type Document) que formata os dados do XML e XSL (Extensible Stylesheet Language), que formata o documento XML noutro tipo de documento para definir os objetos de formatação na apresentação gráfica (por exemplo, XSL-FO para formatação de objetos).

Como evoluções recentes destes formatos o XML Schema, XSD que supre algumas limitações e funcionalidades do DTD e o XSLT (eXtensible Stylesheet Language for Transformation) que facilita transformações de XML em outros documentos.

Uma nova conceituação da Web transforma os usuários de meros consumidores em produtores de informação e de conteúdo, com isto a confusão que já não era pequena se acentua, pois além dos conteúdos (blogs, microblogs, auto publicação, etc.), também no campo da marcação surgem novos recursos como as tags sociais (social tagging) que permitem a marcação dos termos de assuntos realizada diretamente pelo usuário.

Porém, os esforços para a estruturação da Web é um contínuo, como exemplo de ações, destaca-se os projetos DBPedia (Paletta, Mucheroni, 2015).

Foi a emergência da Web 3.0 (Hendler, 2009), ainda que, outros conceitos atinentes a área da ciência da informação como: ontologias, tesouros e vocabulários controlados façam parte deste esforço, a possibilidade de ligar dados em acesso aberto (Open Linked Data) tornou estes esforços profícuos.

Algumas questões podem ser levantadas. O que podemos esperar deste cenário? É possível pensar em esforços para uma Web Inteligente? A Web das coisas pode ter uma ligação com isto tudo?

A nossa resposta é afirmativa para estas três perguntas, porém será imprescindível um esforço para a estruturação dos dados na Web, o que implica em uma discussão ampla sobre a padronização de ambientes, recursos e o uso de metadados que torne possível tal estruturação, bem como, ambientes e recursos mais inteligentes, de forma a corroborar com o futuro do desenvolvimento da Web.

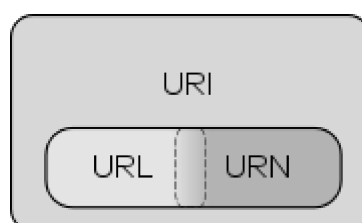
Neste aspecto, o objetivo deste trabalho é o de descrever e refletir sobre a evolução e contribuições da Web. Por meio da análise de alguns dos seus principais conceitos, que a estruturam no que se refere à organização, apresentação, disponibilização e localização de conteúdo, em especial abordando o conceito de metadados, com destaque para os metadados descritivos, apresentando os mais significativos na área da ciência da

informação. Em termos metodológicos, o texto caracteriza-se como descritivo ao elencar e definir os termos relacionados com a web, especialmente os metadados e seu significado para o ambiente web.

2 A WEB 1.0 DOS HIPERLINKS

Uma URN é como o nome de uma pessoa, já a URL é como o seu endereço. A URN define a identidade de um item, enquanto que a URL nos fornece um método para encontrá-lo. A figura 1 ilustra a composição de uma URI, além de comentar na sequência suas características.

Figura 1 – Uma ilustração da composição do URI



Fonte: autores

Uma URN típica pode ser comparada ao sistema International Standard Book Number - ISBN para identificar individualmente os livros. No exemplo: ISBN 0-486-27557-4 (urn: isbn: 0-486-27557-4) é citada, sem equívocos, uma edição específica da obra de Shakespeare "Romeu e Julieta". Para acessar a este objeto e o ler, é necessário obter a sua localização: um endereço URL.

Uma URL para este livro pode ser exemplificada como um caminho de arquivos, como: file:///home/pedro/Desktop/RomeoAndJuliet.pdf, identificando o livro eletrônico salvo no disco de um computador pessoal local. Então o propósito da URN e da URL é o de serem complementares.

A recomendação RFC3305 (Request for Comments), intitulada "Reportagem do planejamento da união W3C/IETF URI interessa grupo: Uniform Resource Identifiers (URIs), Uniform Resource Locator (URLs) e Uniform Resource Names (URNs): Explicações e Recomendações" esboçam o esforço do grupo de trabalho W3C/IETF constituído para normalizar as visões divergentes envolvidas na terminologia das "UR*". Tornou-se a base para um entendimento comum dos padrões e que pode ser referenciada.

As URIs típicas são do tipo: "HTTP", "FTP", "mailto", "URN", "tel", "rtsp", "file" etc., seguido do caractere dois pontos (:) e, por fim, a parte específica do conjunto. A sintaxe e a semântica desta parte específica são determinadas pelas especificações que governam o conjunto, embora a sintaxe URI force todos os conjuntos a aderirem a uma síntese genérica que, além de outras coisas, reserva certos caracteres para propósitos especiais.

Em HTML os elementos básicos que são componentes de um URI devem estar identificados como: o valor de “src”, atributo do elemento “img”, é uma referência URI, como também o valor do comando “href” para os elementos “a” ou “link”.

Os atributos dos comandos em HTML são tags, isto é, podem ser considerados metadados, mas são insuficientes para estabelecer significação semântica. Apesar de descritores, eles não são suficientes para fornecer uma semântica, assim surgiu o XML.

3 A WEB 2.0 DO SOCIAL E DO USUÁRIO

A Web 2.0 surge a partir de ferramentas que possibilitaram ao usuário usar a Web como uma plataforma (Tim O’Reilly, 2005). Entretanto, o usuário nem sempre tem preocupação com a estruturação dos conteúdos e dos dados. Neste sentido, Tim Berners-Lee manifesta a preocupação com a estruturação semântica da Web. A necessidade da Web Semântica que já era clara no início da Web (Berners-Lee, 2001), torna-se imperativa com a sua popularização.

Uma estrutura XML, grosso modo é dada pelo identificador de sistema sucedendo a palavra-chave SYSTEM em uma DTD é uma referência URI sem fragmentos.

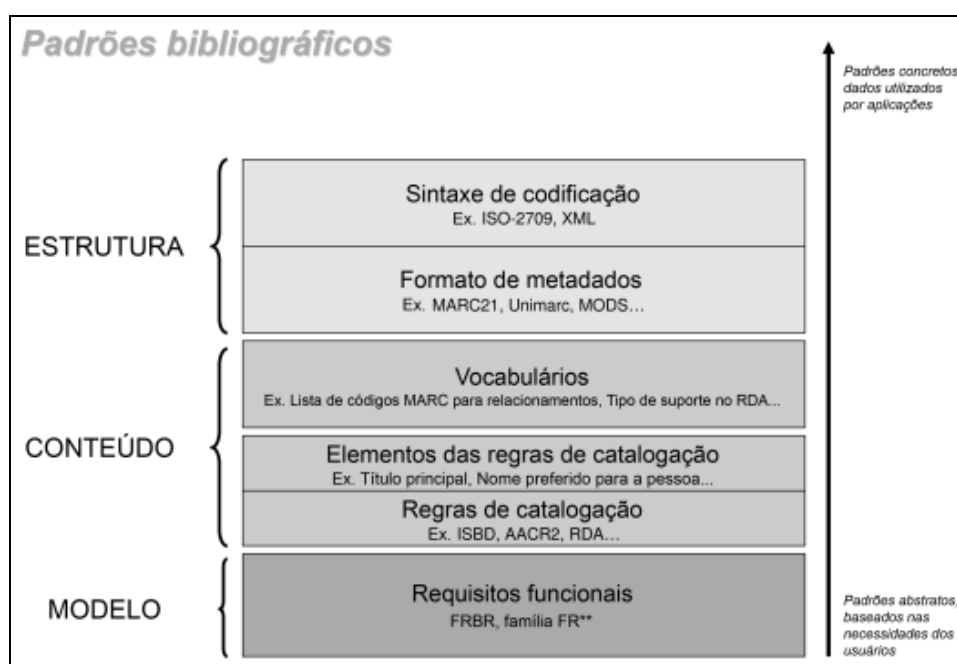
Já, em XSLT, o valor do atributo “href” do elemento/instrução “xsl: import” é uma referência URI, como também o primeiro argumento da função “document ()”.

A grande diferença da Web Semântica que organiza e facilita o intercâmbio semântico dos dados entre homem e máquina, sem os necessários padrões, se origina da Ciência da Informação, onde os documentos poderão ter uma semântica ambígua, mas seus processos de certificação garantem a autoria confiável.

Com esta base na origem conceitual, colaboram-se no surgimento de complementos para a estruturação da linguagem de marcação da Web, os padrões de metadados normalizados, como: DC (Dublin Core), MARC (MAchine-Readable Cataloging), MODS (Metadata Object Description Standard), MADS (Metadata Authority Description Schema), entre outros.

Complementar porque os dados inseridos em ambientes normalizados, isto é, com metadados descritivos, facilitam a busca de documentos e, embora, o usuário não use e também não perceba, são estes dados que certificam e dão encontrabilidade nas buscas realizadas. Assim, prevemos uma maior integração de ambientes interativos na Web com uso cada vez mais intensivo de metadados descritivos, como o mencionado DC, amplamente utilizado, além de outros padrões emergentes e modelagem de relacionamento de dados: FRBR (Requisitos Funcionais para Dados Bibliográficos), FRAD (Requisitos Funcionais para Dados de Autoridade), ambos apoiados por uma nova norma descritiva denominada RDA (Recursos: Descrição e Acesso). A figura 2 ilustra os padrões bibliográficos e as estruturas padrões de seus conjuntos de metadados.

Figura 2 – Metadados de estrutura, modelo e conteúdo.



Fonte: <http://fabricioassumpcao.com/tag/frbr/page/3>

4 METADADOS: CONCEITOS, FUNDAMENTOS DE ESTRUTURA E

MODELO

Metadados, do termo inglês metadata, é uma forma de estruturar e buscar conteúdos dispostos na Internet. Pode ser entendido, também, como o meio de descobrir quais os recursos existentes e como podem ser obtidos e acessados, evitando a ambiguidade dos dados. Baseiam-se em tags contidas nos formatos de marcação (HTML ou XML).

Os metadados são importantes para a identificação, organização e recuperação da informação digital. Sua finalidade é facilitar, globalmente, a localização e recuperação das informações eletrônicas, para os usuários. Neste sentido, utiliza-se dos procedimentos técnicos de indexação e classificação dos conteúdos informacionais, possibilitando a integração de fontes diversificadas e heterogêneas de informação (SHAEFER, 1998).

Mendez Rodriguez (2002) ao conceituar, justificar e fundamentar os metadados comenta que o termo é utilizado como uma palavra descritiva em si mesmo. Entretanto, destaca que há excessiva utilização, com diferentes significados que dependem do contexto de sua aplicação, em especial, no âmbito da biblioteconomia. A definição pode ser vaga se não agregar um qualificador que precise seus objetivos, conteúdo e campo de aplicação.

A origem do termo inicia-se, segundo Howe (2007), nos anos de 1960, cunhado por Jack E. Myers para descrever um conjunto de dados, além da finalidade comercial para designar sua empresa de informática: Metadata Information Partners.

O termo começa a aparecer com frequência na década de 1980, na literatura sobre sistema de gestão de bases de dados, conforme a indicação da base de dados de resumos, o LISA (Library and Information Science Abstracts), para um dos primeiros artigos definindo o termo, no contexto dos sistemas de gestão de bases de dados, como: “informação sobre dados numéricos [...] um sistema de metadados funciona como um equivalente paralelo aos sistemas de dados numéricos de forma que os gestores de sistemas ou outros profissionais aprendam suficiente sobre dados numéricos para ressaltar a probabilidade de validade de seu uso (LISTON; DOLBY, 1982: p. 147.)

Por outro lado, o entendimento sobre metadados estava aplicado à descrição de recursos que documentavam as características das informações contidas nas bases de dados. Entretanto, ainda, nos anos de 1980, na literatura biblioteconômica utilizava-se de expressões como descrições bibliográficas, dados bibliográficos ou catalográficos, ou simplesmente catalogação para denominar o conceito de metadados – dados sobre dados, tanto os recursos catalogados na biblioteca (monografia, mapas, vídeos etc.), como os registros que descreviam estes recursos. A mesma denominação era adotada quando o registro bibliográfico se realizava de forma eletrônica com o formato MARC. Também é considerada uma mudança de concepção na descrição dos recursos de informação, no contexto bibliotecário com a revisão das normas catalográficas. Em especial, as ISBD (Descrição Bibliográfica Normalizada Internacional), cujo capítulo “Arquivo de Computador” passa a denominar-se “Recursos Eletrônicos”, ampliando a tipologia dos arquivos computacionais aos arquivos eletrônicos remotos da internet. Entretanto, as normas atualizadas na nova concepção como o AACR (código de catalogação Anglo-Americano) não mencionam o termo metadados. Porém, “Todas estas circunstâncias serviram para contemporizar a prática da catalogação tradicional com o novo conceito, de tal forma que metadados se converteu em um termo aceito em todas as disciplinas relacionadas à organização da informação e o conhecimento da internet (MENDEZ RODRIGUEZ, 2002, p.33).

Em meados da década de 1990, o termo aparece em comunidades implicadas com a gestão e interoperabilidade de dados geoespacial, e em projetos e manutenção de sistemas de gestão de dados. Para estas comunidades, os metadados se referiam a um conjunto de normas comerciais, junto a uma documentação adicional e outros dados necessários à identificação, representação, interoperabilidade, gestão técnica, funcionamento e uso de dados armazenados em um sistema de informação. Esses componentes normativos de estruturação e definição da informação é o que persiste no significado do termo, até estas primeiras décadas do século 21. Entretanto, a evolução do termo não pode excluir ou omitir sua aplicação no âmbito da biblioteconomia (MENDEZ RODRIGUEZ, 2002).

4.1 CONCEITOS E FUNDAMENTOS

Apesar do enfoque deste artigo estar centrado no contexto da web, em especial da web semântica, ao tratar dos metadados é quase impossível estabelecer uma definição unívoca. Cada área de conhecimento adota o termo, dentro de seu contexto informacional. Para um engenheiro de sistemas, metadados define um nível físico de

informação em relação ao formato do arquivo; para os gestores de bases de dados, tratam dos conteúdos dos objetos de dados contidos na base, seus atributos e relacionamentos.

Ao abordar o termo sob o enfoque da web semântica, em todos os contextos de informação, metadados é um elemento essencial. Em comum, entende-se a sua capacidade de informar sobre os conteúdos que incluem em si mesmos dados descritivos. Assim, da mesma forma que os documentos impressos se autocontemplam alguns dados descritivos (implícitos), como: autor, título, resumo (caso de artigo científico); da mesma maneira as versões digitais incluem dados descritivos. Por exemplo, conteúdos em HTML possuem etiquetas específicas como <TITLE>...</TITLE>, que permitem incluir metadados no cabeçalho do documento web, localizado entre <HEAD>...</HEAD>. Mesmos formatos de imagem como PNG (Portable Network Graphics) ou TIFF (Tagged Image File Format) admitem descrições estruturada de dados. Mensagens de correio eletrônico (e-mails) contêm no cabeçalho da mensagem metadados simples, basicamente, dados de autor, destinatário, assunto e data da correspondência. Segundo Mendez Rodriguez (2002) qualquer entidade ou agrupamento de dados que pretenda dar utilidade ao seu conteúdo precisa incluir informações descritivas sobre seus dados.

Dempsey e Heery (1998, p.149) ao opinarem sobre o futuro da Rede, tratam das distintas interpretações do termo metadados. Situam seus entendimentos no contexto da informação eletrônica distribuída pela web, e os documentos como um objeto. Assim, os metadados são dados vinculados a objetos que liberam seus usuários potenciais (pessoas e máquinas) da necessidade de obter domínios avançados sobre sua existência e características dos conteúdos. “Metadados são conhecimentos que permitem aos usuários, humanos e máquinas, comportarem-se de maneira inteligente. A esta compreensão acrescentou a definição de Tim Berners-Lee (1997) que define metadados como “informação inteligível por computador sobre recursos web e outras coisas”.

As definições e comentários até aqui apresentados destacam três aspectos essenciais dos metadados: a finalidade, o contexto e o interlocutor (homem ou máquina). Estão adensadas pelo próprio W3C, que acrescentam a estes aspectos a função identificadora e descritiva dos metadados, o contexto da Rede, e a possibilidade de interpretação pela máquina. Portanto, o termo metadados se refere a qualquer dado que auxilie na identificação, descrição e localização de recursos na web.

Quanto à aplicação dos metadados, podem ser compreendidas sob a ótica da ciência da informação em dois contextos: função tradicional e atual; e a função emergente e atual com a internet. As aplicações das funções de metadados não são excludentes e nem hierárquicas em termos de prioridades. Apenas ampliam sua importância de forma crescente, conforme o aumento dos conteúdos digitais como objetos de informação.

Quadro 1 – Funções dos Metadados

Tradicional/Atual	Emergente/Atual com a Internet
Identificação e descrição dos recursos; Busca e Recuperação dos recursos; Localização dos recursos;	Autoria e propriedade intelectual sobre os recursos; Atualização dos recursos; Preservação e Conservação; Condições de acesso e uso dos recursos; Valoração dos recursos.

Adaptado: Mendez Rodrigues (2002) e Tammaro; Salarelli (2008)

As funções elencadas no quadro 1 podem ser enfocadas sob o ponto de vista do usuário e, também, do ponto de vista do sistema automatizado. Em termos de sistemas, os metadados facilitam a interoperabilidade e o compartilhamento de dados entre distintas ferramentas de busca. Em termos de usuário facilita detalhes sobre a disponibilidade de acesso ao conteúdo e suas condições de uso.

4.2 MODELOS OU TIPOLOGIA DE METADADOS

Certamente, existem distintos tipos de metadados, na medida em que há, também, uma diversidade de usuários, com diversificadas necessidades e variados tipos de informações. Sintetizando interpretações propostas por Caplan (2003), Greenberg (2000), e Lagoze (1996), apresentam-se as conotações distintas da função desempenhada pelos metadados. São estes:

- Metadados administrativos: informações sobre o recurso em si (a sua criação, modificação, relacionamento com outros registros, etc.). Estes metadados podem incluir: número do registro data de sua criação e modificação, a identificação do criador e revisor; linguagem; notas; relação com recurso (s) outro (s).
- Metadados descritivos: informações sobre as propriedades físicas e intelectuais ou o conteúdo de um recurso digital, que pode conter: título, criador, data, editora; identificadores únicos (ISBN, ISSN, etc.); links dinâmicos (URI; URL, etc.); resumo; nota descritiva; revisão; formato; versão de idioma, etc.
- Metadados analítica: informações que permitem explorar a melhora no acesso ao conteúdo do recurso, que pode incluir: termos de assunto controlado, abstrato; tabela de conteúdo, códigos derivados de sistemas de classificação; outros elementos de importância local; links para outros conteúdos relacionados.
- Metadados de Gestão de Direitos: informações sobre restrições / permissões (legal; financeiros, etc.) sobre o acesso ou uso de itens digitais ou objetos.
- Metadados técnicos: informações sobre hardware ou software usado na conversão de um item de informações para um formato digital, ou no armazenamento, visualização, entre outros.
- Metadados de preservação: informações sobre a condição física de um item / objeto e as ações (recarregamento, migração, etc.) realizada para preservar e gerenciar item físico e digital / objetos / coleções.
- Metadados estruturais: informações sobre os tipos, versões, relacionamentos e outras características de itens digitais ou objetos, e/ou suas partes componentes.
- Metadados do uso: informações sobre o nível e tipo de uso de recursos de informação.
- Outros metadados: elementos de especial aplicação local, regional, nacional e/ou requisitos organizacionais.

Como o tema não se esgota com o presente texto, descrevem-se de forma simplificada modelos relativos aos padrões descritivos, mais citados na literatura da área da ciência da informação, como ilustração da sua variedade e características, no contexto da web. Os modelos empregados são:

- MARC (Machine Readable Cataloging), formato para programa computacional que interprete os dados de um registro bibliográfico, é adotado no ambiente das bibliotecas. A padronização dos dados bibliográficos, utilizando o formato, assegura a integridade do catálogo da biblioteca, no que se refere ao

armazenamento e recuperação das informações. Padrão mantido pela Library of Congress dos Estados Unidos.

- TEI (Text Encoding Initiative) é um consórcio internacional, criado em 1987 para desenvolver diretrizes para a codificação de textos legíveis por máquina nas áreas das ciências humanas e sociais. Estruturalmente, é composto de uma linguagem de codificação extensa, orientada à codificação de documentos muito complexos. Os conceitos básicos são simples e a codificação utilizada não é mais difícil ou complexa do que as do HTML (TEI, 2015).
- EAD (Encoded Archive Description), esquema baseado na SGML, é mantido pela Society of American Archivists e pela Library of Congress. Trata-se de padrão não-proprietário para codificar conteúdo na web. O seu objetivo comum é fornecer descrição detalhada de conteúdos organizados e de coleções de materiais de arquivo (EAD, 2015).
- ONIX (Online Information Exchange) é um formato desenvolvido pelos editores para distribuir informações eletrônicas sobre suas publicações a livreiros, agregadores, distribuidores de livros, outros editores, e qualquer outra pessoa envolvida na comercialização de livros. O padrão define uma lista de elementos pré-definidos sobre um recurso (livro), para se ter a certeza que se referem à mesma coisa. Alguns desses elementos, tais como: ISBN, autor e título, são necessários; outras como resenhas de livros e imagem da capa são facultativos. O formato adota o XML entre outras razões por: ser otimizado para a criação de documentos complexos e transmissão e troca de dados entre computadores. Os textos em XML são legíveis por máquina (NETREAD, 2015).
- MODS (Metadata Object and Description Schema), elaborado pela Library of Congress, é um esquema de descrição bibliográfica, baseado em XML, e concebido como uma espécie de ponte entre a complexidade do formato MARC, utilizado pelas bibliotecas, e a extrema simplicidade do formato de metadados Dublin Core (MODS, 2015).
- METS (Metadata Encoding and Transmission Standard), elaborado também pela Library of Congress, é um esquema baseado em XML, para a codificação de metadados sobre objetos descritivos, administrativos e estruturais de recursos contidos em uma biblioteca digital. O padrão consiste de sete seções: 1) cabeçalho, contendo metadados que descrevem o documento METS (como criador, editor, etc.); 2) metadados descritivos que pode apontar para metadados externos ao documento METS (por exemplo, um registro MARC em um catálogo mantido em um servidor web), ou contém metadados descritivos interno ao ambiente da biblioteca digital, ou ambos; 3) metadados administrativo que fornecem informações sobre como os arquivos criados e armazenados, os direitos autorais existentes, metadados sobre o objeto da qual deriva o conteúdo, etc.; 4) seção de arquivo, provendo as listas de todos os arquivos que contém o conteúdo do objeto digital; 5) mapa da estrutura que é o coração de um documento METS. Ele descreve uma estrutura hierárquica para o objeto da biblioteca digital e vincula os elementos da estrutura dos arquivos de conteúdo e metadados que pertencem a cada elemento; 6) seção das ligações estruturais, permite que criadores METS registrem a existência de hiperlinks entre os nós existentes na hierarquia delineadas do mapa da estrutura (Isto é valioso para arquivar sites, no METS); 7) seção de comportamento, pode ser usada para associar comportamentos executáveis com objetos de conteúdo no METS (METS, 2011).

- DC (Dublin Core Metadata Element Set), uma das importantes iniciativas de sucesso, apoiada por iniciativa de associações de bibliotecários norte-americanos, e pela OCLC (Online Computer Library Center). O padrão considera os documentos como objetos e integra todos os possíveis tipos de materiais e seu tratamento técnico, sejam bases de dados, imagens digitais, banco de imagens, textos, textos eletrônicos, vídeos e películas em formato digital e não digital objeto multimídia, gravações sonoras em formato digital e não digital. Desta forma, visa integrar os diversos catálogos em um domínio de acessos cruzados e múltiplos através de conexões hipertextuais. Saliente-se que todos os seus elementos são opcionais; podem ser exibidos aleatoriamente (em qualquer ordem); além de extensível, podem coexistir com outros metadados.

Acrescente-se que os quinze elementos "Dublin Core", descrito no padrão é parte de um conjunto mais amplo de vocabulários de metadados e de especificações técnicas mantido pela Dublin Core Metadata Initiative (DCMI).

Ademais é simples e autoexplicativo, com o próprio autor descrevendo seu conteúdo (MENDEZ RODRIGUEZ, 2002). Indicação dos elementos no quadro 2.

Quadro 2 – Elementos da estrutura do DC

Title	Contributors	Language
Creator	Date	Relation
Subject	Type	Coverage
Description	Format	Rights
Publisher	Identifier	management
	Source	
	Fonte: DCMI	

O conjunto completo dos vocabulários, DCMI Metadata Terms [DCMI-termos], também inclui conjuntos de classes de recursos (incluindo o tipo Vocabulário DCMI [DCMI-TYPE]), esquemas de codificação de vocabulário e sintaxe de esquemas para codificação. Os termos no vocabulário DCMI destinam-se a ser utilizado em combinação com outros termos, a partir de vocabulários compatíveis no contexto dos perfis de aplicação, e na base do DCMI Abstract Model [DCAM] (DCMI, 2015).

Outros padrões que têm emergido citados por Tamara e Selarei (2008), são:

- DOI (Documento Objeto Identifier) é um identificador destinado à gestão de objetos digitais protegidos pelo direito autoral. Desenvolvido pela Association of American Publisher, e mantido pela International DOI Foundation. O DOI é um identificador persistente para conteúdo digital, compatível com URN, mas somente isso é, sobretudo uma entrada para uma variedade de serviços associados ao uso do conteúdo. Significa que para o acesso aos variados serviços possíveis para qualquer objeto digital, deve ser possível o acesso as várias fontes de informação sobre este objeto. Para um conteúdo ser útil necessita de uma estrutura estável.
- FRBR (Functional Requirements for the Bibliographic Record), desenvolvido pela IFLA (International Federation of Library Associations and Institutions), constitui-se de um modelo conceitual orientado à harmonização da descrição de diversos objetos digitais e as relações entre eles. Caracteriza-se em quatro estados no qual o recurso informacional pode estar situado. O autor ou criador do conteúdo (obra), que é um conceito abstrato, expressa o conteúdo de maneira

física ou digital. Toda obra pode ter muitas expressões. Toda expressão pode estar inserida em uma ou muitas manifestações que pode ser a produção do conteúdo em uma imagem, texto, etc. Toda manifestação é reproduzida em certo número de itens (exemplares).

- FRAD (Functional Requirements for Authority Data) é também um modelo conceitual proposto pela IFLA, com a finalidade de proporcionar parâmetros de análise dos tipos de dados de autoridade e do intercâmbio internacional destes dados. Basicamente, o modelo foi projetado para proporcionar indicação referencial definida e estruturada; relacionar os dados que relacionam os criadores de registros de autoridade com as necessidades dos usuários destes dados; auxiliar na valoração das possibilidades do intercâmbio internacional; e utilização dos dados de autoridade, tanto pelo setor bibliotecário, como por outros setores (PATTON, 2009).

O modelo FRAD conta com uma ampla lista de atributos possíveis para a entidade pessoa física, atributos encontrados em dados de autoridade. Tais atributos são identificados como: Data relativa à pessoa; Título da pessoa; Gênero; Lugar de nascimento; Lugar de morte; País; Lugar de residência; Afiliação; Endereço; Língua da pessoa; Campo de atividade; Profissão/ocupação; Biografia/história; outras informações relativas à pessoa.

Uma das aplicações, na web, dos conceitos preconizados no FRAD encontra-se no projeto VIAF (Virtual International Authority File), um serviço internacional projetado para acessar sistemas de autoridades das principais agências bibliográficas do mundo (limitado nesta etapa inicial aos nomes de pessoas). É um esforço conjunto da Library of Congress, Deutsche Nationalbibliothek, Bibliothèque Nationale de France e OCLC (Online Computer Library Center), em cooperação com um número crescente de outras bibliotecas nacionais. O VIAF foi pensado como um dos componentes essenciais da Web Semântica, e que permite apresentar os nomes de pessoas em ordem alfabética e língua de cada usuário (OCLC, 2015).

Os modelos FRBR e FRAD promovem mudanças de perspectivas de como os dados são analisados quanto sua utilidade para o seu usuário. Os modelos também assumem a perspectiva da navegação em grandes catálogos e bases de dados (OLIVER, 2011). Outro aspecto é a expectativa de que o padrão de metadados mais difundido, o Dublin Core, acrescido do DOI, e auxiliados pelo modelo FRBR e FRAD, sejam empregados de modo sinérgico no tratamento da descrição dos recursos digitais.

Em um artigo de Christian Bizer (2006) criador do projeto DBPedia, descreve Linked Data como o uso da Web para criar links digitados entre dados de diferentes fontes (recuperáveis na Web, bastando o XML correspondente) e tecnicamente refere-se a dados publicados na Web, de tal forma que é legível por máquina, onde o seu significado seja explicitamente definido, estando ligado a outros conjuntos de dados externos, e por sua vez podem estar ligados a partir de conjuntos de dados externos (links que descrevam estes dados), porém estes dados carecem de uma ligação confiável de autores e assuntos.

Assim, Linked Data tornou-se uma tecnologia emergente na Web Semântica por que tem como base duas tecnologias fundamentais da Web: Uniform Resource Identifiers (URIs) (Berners-Lee, 2006) e Hypertext Transfer Protocol (HTTP) (Fielding et al, 1999) agora atualizadas.

Neste artigo Berners-Lee simplifica colocando quatro pontos fundamentais para uma boa descrição na Web Semântica: 1- Uso de URIs como nome para as coisas, 2. Uso de HTTP e URIs de modo que as pessoas podem ser vistas nos nomes, 3. Quando alguém procura um URI, fornecer informações úteis, usando os padrões (RDF, SPARQL) e 4. Incluir links para outros URIs, para que eles possam descobrir mais coisas.

Esclareceu-se neste trabalho que todas estas tarefas, em especial, a segunda e a terceira, não fazem isto de maneira precisa sem o uso de metadados, portanto, é uma tecnologia essencial.

5 CONCLUSÃO

O trabalho descreve a evolução e contribuições da Web, repassando os conceitos fundamentais quanto à organização da Web Semântica, e procura mostrar a essencialidade dos metadados, e de modo particular e enfático, a importância dos conceitos de metadados descritivos que advém da Ciência da Informação e que devem ser incorporadas as novas construções na Web Semântica, a exemplo do que foi feito no projeto VIAF (The Virtual International Authority File), em relação à autoria.

Também foi necessário precisar os termos relacionados com a Web e a Web Semântica com a finalidade de descrever de modo simples e didático o que são os descritores necessários para uma boa documentação e recuperação de dados na Web, e como uma linguagem de marcação com finalidade semântica, o XML e sua gramática em RDF, auxiliam este processo.

As diversas tecnologias de descrição de metadados advindas dos estudos e grupos de padronização da Ciência da Informação, ao longo da história e também agora auxiliaram estes desenvolvimentos, casos bastante claros são o MARC (Machine Readable Cataloging) e DC (Dublin Core), agora o FRBR (Functional Requirements for Bibliographic Records).

Também os padrões emergentes de certificação de documentos na Web como o DOI (Document Object Identifier), destinado à gestão de objetos digitais orientados ao direito autoral, desenvolvido pela Association of American Publisher, o FRBR (Functional Requirements for the Bibliographic Record), desenvolvido pela IFLA (International Federation of Library Associations and Institutions), que descreve diversos objetos digitais e as relações entre eles, e também o FRAD (Functional Requirements for Authority Data), igualmente proposto pela IFLA, proporcionam parâmetros de análise dos tipos de dados de autoridade e do intercâmbio internacional destes dados.

Estes modelos não competem, mas auxiliam o processo de identificação e certificação de dados na Web Semântica, havendo já modelos híbridos com o XML, como são os casos do METS, elaborado também pela Library of Congress, que é um esquema baseado em XML e MODS (Metadata Object and Description Schema), feito também pela Library of Congress, que é um esquema de descrição bibliográfica, baseado em

XML, mas como uma ponte entre a complexidade do formato MARC e simplicidade do formato Dublin Core, usando XML.

Este trabalho teve por objetivo demonstrar a complementariedade dos conceitos desenvolvimentos pelos métodos tradicionais descritivos oriundos da Biblioteconomia e Ciência da Informação e sua integração com as descrições vindas da Web Semântica e que também trazem necessidade de padronização e organização, tecnologias historicamente desenvolvidas na Ciência da Informação.

O trabalho ainda demanda tempo e esforço em pesquisa, mas o complexo mundo informacional que a Web criou pode e deverá usar cada vez mais conceitos de organização da informação e do conhecimento que são apresentados como uma contribuição da Ciência da Informação para o desenvolvimento da área.

REFERÊNCIAS

ABADI, D. J. et al. Scalable Semantic Web Data Management Using Vertical Partitioning. Proc. 33rd Conf. Very Large Data Bases, VLDB Endowment, 2007, p. 411-422.

ASSUNÇÃO, F. Padrões bibliográficos. Março de 2012 Disponíveis em: <http://fabricioassumpcao.com/tag/frbr/page/3>. Acesso: 12/12/2015.

BERNERS-LEE, T. Information Management: a proposal. 1989. Disponível em: <http://goo.gl/oovRMi>. Acesso em: 4-11-2014.

BERNERS-LEE, T. Metadata architecture: document, metadata and links. Design Issues, W3C, 6 de janeiro de 1997. Rev. 27/08/2009. Disponível em: <http://www.w3.org/DesignIssues/Metadata>. Acesso em 26/11/2015.

BERNERS-LEE, T. Linked Data - Design Issues. Retrieved July 23, <http://www.w3.org/DesignIssues/LinkedData.html> , 2006.

BIZER, Christian Linked Data - The Story So Far, Freie Universität Berlin, Germany, 2006.

BIZER, C.; Heath, T. and Berners-Lee, T. Linked data – the story so far, International Journal on Semantic Web & Information Systems 5(3), p. 1–22, 2009.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American Magazine, May, 2001. Disponível em: <http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html>. Acesso em 19-11-2014.

CAPLAN, P. Metadata Fundamentals for All Librarians. USA: ALA, 2003.

DCMI. Dublin Core Metadata Element Set, Version 1.1: introduction. Disponível em: <http://www.dublincore.org/documents/dces/>. Acesso em: 11/12/2015.

DEMPSEY, L.; HEERY, R. Metadata: a current view of practice and issues. *Journal of Documentation*, vol. 54, n.2, p.145-172, Mar. 1998.

EAD. General information. Disponível em: <http://www.loc.gov/ead/index.html>. Acesso em 12/12/2015.

FIELDING ET al., HTTP 1.0 Hypertext Transfer Protocol, W3C Consortium, 1999, Disponível em: <https://www.w3.org/Protocols/rfc2616/rfc2616.txt> , Acesso: dez. 2015.

GREENBERG, J. (ed.). *Metadata and Organizing Educational Resource on the internet*. New York: Harworth Information Press, 2000.

HENDLER, J. Web 3.0 Emerging. *Computer*, v.42, n.1, p.111- 113, jan. 2009. Disponível em: <http://dl.acm.org/citation.cfm?id=1512177>. Acesso em: 24-11-2014.

HENDLER, J. Web 3.0: The Dawn of Semantic Search, *IEEE Computer*, jan. 2010. Disponível em: <http://goo.gl/tcA2oz>. Acesso em: 25-11-2014.

HOWE, D. FOLDOC: Free Online Dictionary of Computing. Rev. 25/07/2007. Disponível em: <http://foldoc.org/>. Acesso em 12/12/2015.

LAGOZE, C. The Warwick Framework: A Container Architecture for Diverse Sets of Metadata. *D-Lib Magazine*, Jul./Aug. 1996. Disponível em: <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>. Acesso em: 12/10/2015.

LISTON, M. D.; DOLBY, L. J. Metadata systems for integrated access to numeric data files. *Drexel Library Quarterly*, vol. 18, n. 3-4, p. 147-160, 1982.

MARKOFF, J. *Entrepreneurs See a Web Guided by Common Sense*. 2006. Disponível em: <http://goo.gl/vzHt3O>. Acesso em: 18-11-2014.

MENDEZ RODRIGUEZ, E. *Metadados y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales*. Gijon, Espanha: Ediciones Trea, c2002.

METS. METS: An Overview & Tutorial. Setembro de 2011. Disponível em: <http://www.loc.gov/standards/mets/>. Acesso em: 12/12/2015.

MODS. MODS: Uses and Features. May 2015. Disponível em: <http://www.loc.gov/standards/mods/>. Acesso em 12/12/2015.

NETREAD. ONIX. Disponível em: <http://www.netread.com/onix/>. Acesso em: 12/12/2015.

OCLC. VIAF: Virtual International Authority File. Disponível em: <http://www.oclc.org/viaf/en.html>. Acesso em: 11/11/2015.

OLIVER, C. *Introdução à RDA: um guia básico*. Brasília: Briquet de Lemos/Livros, 2011.

PALETTA, F. C.; MUCHERONI, M. L. O Desenvolvimento da Web 3.0: Liked Data e DBPedia. *PRISMA.COM*, n.25, p. 73-90, 2014.

PATTON, G. E (ed.). Functional requirements of authority data: a conceptual model: final report, December 2008. München: K. G. Saur, 2009.

REITZ, J. M. ODLIS: Online Dictionary for Library and Information Science. 2007. Disponível em: http://www.abc-clio.com/ODLIS/odlis_A.aspx. Acesso em: 20/10/2010.

RFC 3160. Relatório do grupo IETF. Disponível em: <https://tools.ietf.org/html/rfc3160>. Acesso em: 15/12/2015

SCHAEFER, M. T. Demystifying metadata: initiatives for web document description. Information Retrieval & Library Automation, vol. 33, n. 11, 1998.

SCOTT, Michael L. *Programming Language Pragmatics*. San Francisco: Morgan Kaufmann/Academic Press, 2006. 875 p. 15.

TAMMARO, A. M.; SALARELLI, A. A biblioteca digital. Brasília: Briquet de Lemos/Livros, 2008.

TEI. <Text Encoding Initiative>. Disponível em: <http://www.tei-c.org/index.xml>. Acesso em 12/12/2015.