

XIV Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB 2013)
GT 8: Informação e Tecnologia

Comunicação Oral

**O USO DO *RESOURCE DESCRIPTION FRAMEWORK* NA ORGANIZAÇÃO DA
INFORMAÇÃO**

Maria Elisabete Catarino – UEL
Brígida Maria Nogueira Cervantes – UEL
Terezinha Batista de Souza – UEL

Resumo

Apresenta algumas aplicações, existentes no âmbito da Ciência da Informação (CI), de uso do *Resource Description Framework* (RDF) para a Organização da Informação. O RDF é o modelo de descrição para recursos da Web proposto pelo *World Wide Web Consortium* (W3C). Com base neste modelo de descrição e outras tecnologias relacionadas aos vocabulários, inferências, busca de informação, ontologias, entre outras, o W3C pretende tornar possível a Web Semântica, na qual os todos dados disponíveis na internet, sejam eles armazenados em sites ou bases de dados de variados tipos (XML, relacionais, etc), estejam legíveis por máquinas e, principalmente, interligados (*Linked Data*), criando o que se chama de Web de Dados. Por outro lado, na CI existem os procedimentos tradicionais de tratamento da informação. O objetivo deste trabalho é identificar ações que estão sendo adotadas pela Ciência da Informação no sentido de tornar os dados (bibliográficos e/ou arquivísticos) legíveis por máquina e lincados. Como resultado de uma pesquisa bibliográfica desenvolvida para cumprir o objetivo citado atrás, alguns exemplos de aplicações em RDF são apresentados. Destaca-se que as tradicionais práticas da Organização da Informação, tanto no que tange a representação descritiva quanto a temática, estão sendo aplicadas de forma a tornar os dados organizados conforme os propósitos da Web Semântica e que, na formação do profissional da informação, é imprescindível incluir conteúdos relacionados à organização da informação na Web.

Palavras-chave: *Resource Description Framework*. Organização da Informação. Representação da Informação. Sistemas de Organização do Conhecimento.

Abstract

Presents some applications existing within the Information Science (CI) for the use of the *Resource Description Framework* (RDF) in the Organization of Information. The RDF is a model for web resources description, proposed by the *World Wide Web Consortium* (W3C). Based on this description model and an other related technologies; such as vocabularies, inference, information search, ontologies, among others; the W3C aims to make possible the Semantic Web, in which all available data on the Internet stored in databases or websites of various types (XML, relational, etc.), are machine-readable and mainly interconnected (linked data) creating what is called Web of Data. On the other hand, in the IC, there standards procedures for processing information. The objective of this work is to identify actions that are being taken by the Information Science for make data machine-readable and linked. As a result of search literature developed to fulfill the objective mentioned above, some examples of applications are presented in RDF. It is noteworthy that the traditional practices of Information Organization, both in terms of descriptive representation as the thematic representation, are being applied to make the data organized according to the purpose of the

Semantic Web The training of information professionals is essential to include related content the organization of information on the Web.

Keywords: Resource Description Framework. Information Organization. Information Representation. Knowledge Organization System.

1 INTRODUÇÃO

A Ciência da Informação (CI) tem como objeto de estudo a informação e atua desenvolvendo pesquisas relacionadas à sua origem, coleta, organização, armazenamento, recuperação, transmissão, além de sua utilização. No tocante a informação nas suas mais variadas formas e suportes, torna-se indiscutível a importância da internet neste contexto.

Desde o seu surgimento, na década de 1960, a internet tem evoluído. O *World Wide Web Consortium* (W3C) apoia iniciativas para o desenvolvimento de normas, recomendações e tecnologias que deem suporte à Web Semântica, um projeto que pretende embutir inteligência aos dados disponíveis na web. Para tanto, é necessário que se padronizem tecnologias, linguagens e metadados descritivos. Dentre as iniciativas do W3C há a proposta de um modelo de descrição de recursos da Web, o *Resource Description Framework* (RDF) que é fundamental para a Web Semântica.

Neste sentido, pode-se inferir que as práticas da CI estão relacionadas aos propósitos do W3C, já que a base para a Web Semântica é um modelo de descrição no qual são aplicados metadados descritivos, inclusive os bibliográficos e arquivísticos, tais como *Dublin Core* (DC), *Resource Description and Access* (RDA) e *Encoded Archival Description* (EAD). Um dos pressupostos da linha de pesquisa na qual o estudo foi realizado é de que as tradicionais atividades de representação da informação são aplicáveis à organização da informação da web. Para averiguar esta ideia desenvolveu-se uma pesquisa intitulada “Contribuições da Ciência da Informação para a Web Semântica” da qual será apresentada, neste artigo, parte dos resultados referentes ao objetivo: identificar ações que estão sendo adotadas pela Ciência da Informação no sentido de tornar os dados (bibliográficos e/ou arquivísticos) legíveis por máquina.

Para atingir ao objetivo proposto elegeu-se a pesquisa descritiva como procedimento mais adequado às necessidades e características deste estudo, de vez que predominou a descrição das características do fenômeno em foco. Como métodos foram empregadas a pesquisa bibliográfica e a pesquisa documental. A pesquisa bibliográfica foi realizada na literatura da área de Ciência da Informação para identificar aplicações que já estão sendo implementadas em RDF na organização da informação bibliográfica e/ou arquivística. A

pesquisa documental foi realizada no site do W3C para informações complementares sobre normas, tecnologias e recomendações da web semântica.

Nas próximas seções serão apresentados os conceitos relacionados à Web Semântica bem como as aplicações identificadas e as considerações finais.

2 WEB SEMÂNTICA

O conceito da Web Semântica foi referido em 2001 na publicação de um artigo na revista *Scientific American* por Tim Berners-Lee, James Hendler e Ora Lassila. Os autores explicam que não é uma Web separada, mas uma extensão da atual, na qual a informação é dada com um significado bem definido, permitindo melhor interação entre computadores e pessoas (BERNERS-LEE; HENDLER; LASSILA, 2001). Trata-se de uma nova geração da Web, projeto da W3C cujo intento é desenvolver tecnologias, linguagens, padrões e recomendações que tornem a informação legível pelas máquinas.

Segundo Berners-Lee, Hendler e Lassila (2001), os computadores necessitam ter acesso a coleções estruturadas de informações (dados e metadados) e de conjunto de regras de inferência que ajudem no processo de dedução automática para que seja administrado o raciocínio automatizado.

A Web Semântica é composta de uma filosofia, um conjunto de princípios para *design*, grupos de trabalho colaborativos e uma variedade de tecnologias necessárias para a implantação destes princípios e propostas dos grupos de trabalho. Alguns elementos são expressos em especificações formais que incluem o modelo *Resource Description Framework* (RDF), formatos de intercâmbio de dados (como por exemplo: RDF/XML, N3, *Turtle*, *N-Triples*), e notações tais como *RDF Schema* (RDFS) e a *Web Ontology Language* (OWL), tudo com a intenção de prover uma descrição formal dos recursos, conceitos, termos e relacionamentos num específico domínio do conhecimento (W3C, 2011).

Adicionalmente à Web de Documentos (a primeira geração da Web), a W3C está apoiando o desenvolvimento de tecnologias para dar suporte à Web de Dados (dados contidos em bases na Web). O objetivo é implementar, divulgar e promover linguagens, normas e recomendações que permitam que os computadores possam, a partir de dados dispersos em diversos sites e bases da Web, compor interações na rede e gerar serviços e produtos. O termo Web Semântica representa uma Web de dados lincados (*linked data*). As tecnologias recomendadas pela W3C permitirão que as pessoas e instituições criem armazém de dados,

construam vocabulários e escrevam regras para a manipulação dos dados (W3C, 2010).

Os conceitos básicos que compõem a ideia da Web Semântica e que são a base da Web de Dados (Dados Lincados, Vocabulários, Busca, Inferência e Aplicações Verticais) serão apresentados a seguir.

O próprio Berners-Lee, um dos idealizadores da Web Semântica, considerou que o nome mais adequado para a Web Semântica seria Web de Dados (*Data Web*). (KING, 2007). O intuito é que os dados da Web estejam disponíveis em formatos padrão e passíveis de serem gerenciados por ferramentas da Web Semântica.

Conforme W3C (2010), para tornar a Web de Dados uma realidade é importante ter uma grande quantidade de dados disponível em um formato padrão, acessível e gerenciável por ferramentas da Web Semântica. Além do acesso aos dados da Web, é necessário, ainda, criar relações entre os dados. Dados Lincados é o nome que se dá às coleções de dados relacionados na Web.

Para a publicação e vinculação dos dados, é preciso adotar o modelo recomendado pela W3C, o *Resource Description Framework* (RDF)¹, que é o alicerce da Web de Dados. As tecnologias disponíveis devem ter, como base, o modelo RDF para possibilitar a conversão e/ou o acesso aos dados existentes em bases de dados.

Os vocabulários na Web Semântica definem conceitos e seus relacionamentos. São úteis para disponibilizar termos que podem ser utilizados numa aplicação específica, bem como definir as possíveis relações entre os termos e as limitações no uso destes. Os vocabulários podem também ser denominados “Ontologias”. Segundo o W3C, na prática, não existe uma recomendação para o uso de um ou outro conceito, mas, comumente, a tendência é usar o conceito “Vocabulário” para coleções de termos mais simples e “Ontologia” para as coleções mais complexas².

No contexto da Web Semântica, “Busca” (*Query*) refere-se às tecnologias e aos protocolos para a recuperação da informação na Web de Dados (W3C, 2010). Assim como as bases de dados relacionais ou em XML necessitam de linguagens específicas de consulta (SQL e XQuery, respectivamente), a Web de Dados, com seus dados formatados em RDF, necessita de uma linguagem de busca própria, no caso, a SPARQL. Esta linguagem permite

¹ RDF – Modelo padrão para intercâmbio de dados na Web (RDF WORKING GROUP, 2004).

² O conceito de complexidade usado aqui é no sentido de grau de formalidade, como por exemplo no âmbito das ontologias que, segundo a tipologia de Uschold e Gruninger (1996), vão desde as Altamente Informais (as mais simples), até as Rigorosamente Formais (as mais complexas) (apud CATARINO, 2009, p. 71).

fazer consultas e receber os resultados em *HyperText Transfer Protocol* (http) ou *Simple Object Access Protocol* (SOAP).

Em Lógica inferência é definida como o processo de concluir uma afirmação a partir de outras afirmações. (INFERÊNCIA, 2011). Na Web Semântica, os dados são modelados como um conjunto de relações entre os recursos da Web. Neste contexto, “Inferência” refere-se à descoberta de novas relações entre os recursos web com base nos dados e em algumas informações adicionais sob a forma de um vocabulário ou conjunto de regras (W3C, 2010).

Aplicações Verticais (*Vertical Applications*) é uma expressão usada pela W3C (2010) para se referir às aplicações genéricas, desenvolvidas por comunidades específicas que exploram as tecnologias do W3C. Algumas dessas áreas de aplicação podem formar grupos de trabalho no W3C, colaborando com os demais membros do consórcio, para explorar novas possibilidades para a Web Semântica. Estes grupos trazem importantíssimas contribuições para o aperfeiçoamento das recomendações, tecnologias e padrões desenvolvidos pelo W3C.

A seguir será descrito o RDF, o alicerce da Web Semântica, no qual se pressupõem que, as práticas da CI estejam imbricadas.

3 RDF

O RDF pode ser considerado um alicerce da Web Semântica e que está relacionado com a Representação Descritiva dos recursos da Web. Segundo Allemang e Hendler (2011, p.27) o RDFS e o OWL são as linguagens básicas de representação da Web Semântica, e têm como fundamento o RDF.

RDF é um modelo para a descrição de recursos. Allemang e Hendler (2011, p.49) o definem como um sistema de modelagem de dados que, em comparação a outros, perde na compacidade, porém ganha em flexibilidade. Segundo Manola e Miller (2004) RDF é uma forma de representação dos metadados dos recursos Web, tais como título, autor, data de modificação de uma página da Web, *copyright* e informações sobre licenças de documentos Web ou informações sobre o compartilhamento.

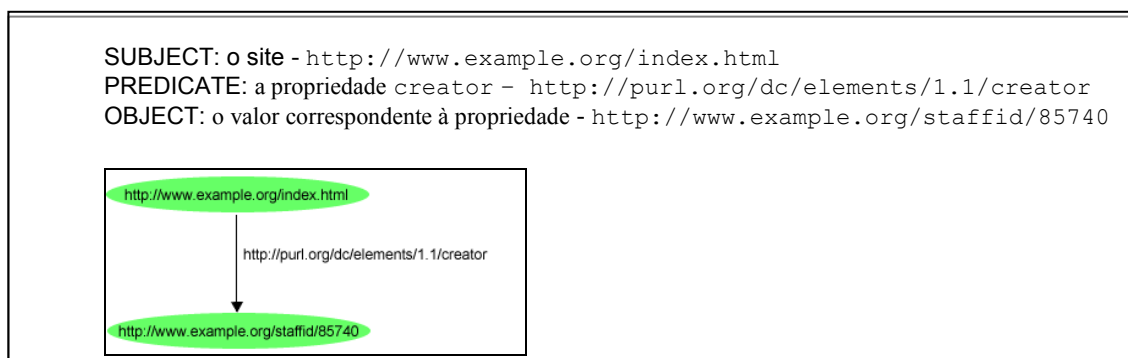
Neste modelo de descrição de dados existe uma estrutura bastante simples. Todas as coisas na Web, ou seja, os recursos Web podem ser descritos em triplos: *Subject*, *Predicate*, *Object* (ALLEMANG; HENDLER, 2011, p.49). Ainda neste sentido, Manola e Miller (2004) dizem que o RDF é baseado na identificação de recursos, por meio de identificadores da Web,

denominados *Uniform Resource Identifiers* (URIs), e na descrição dos recursos, em termos de propriedades e seus valores correspondentes.

Esta estrutura permite uma forma bastante simples de mesclagem de dados (*merging data*). Segundo Allemberg e Hendler (2011) não é necessário estruturar os elementos de metadados em forma de tabelas/colunas para poder mesclar ou recuperar os dados. A mescla dos dados é transformada, a partir dos triplos, numa simples questão de reunião de todas as declarações³, oriundas de diversas fontes num único lugar.

Na figura a seguir apresenta-se um exemplo de descrição de um recurso em RDF:

Figura 1: Uma declaração em RDF.



Fonte: Manola e Miller (2004).

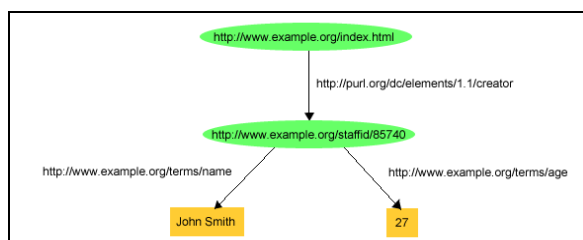
Esta declaração seria representada, em RDF/XML, indicando a versão XML usada, os *namespaces* utilizados (no caso: RDF, DC e EXTERMS) e na sequência a descrição propriamente dita (MANOLA; MILLER, 2004).

```
<?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:ext="http://www.example.org/terms/">
  <rdf:Description rdf:about="http://www.example.org/index.html">
    <dc:creator rdf:resource="http://www.example.org/staffid/85740"/>
  </rdf:Description>
</rdf:RDF>
```

³ Declaração pode ser definida como sendo uma parte dos metadados que descreve um recurso na Web. Composta de um “*subject*”, um “*predicate*” e um “*object*”, ou seja, de um Recurso, uma Propriedade e seu Valor correspondente: “*Moby Dick / has author / Herman Melville*” é uma declaração. Na proposta da Web Semântica, não há um registro de metadados, mas um grupo de declarações relacionadas. As declarações relacionadas expressam a mesma descrição que um registro de metadados expressaria em outros sistemas (W3C, 2011).

O exemplo acima representa uma declaração simples, um recurso e suas propriedades. Contudo, as descrições no modelo RDF podem ser mais complexas. Por exemplo, um valor (OBJECT) poderá ser ele próprio um SUBJECT com suas propriedades (PREDICATE) e seus respectivos valores, conforme ilustra a Figura 2.

Figura 2: Descrição composta.



Fonte: Manola e Miller (2004).

O objetivo deste texto é apresentar a ideia do RDF, com seus princípios básicos. No entanto, ainda existem muitas outras questões importantes que permeiam a descrição por meio do RDF. Uma dessas questões é o fato de um valor poder ser representado tanto por um URI (no exemplo acima <http://www.example.org/staffid/85740> - URI referente ao *CREATOR* da página, por exemplo: John Smith) quanto por um *string*⁴ (no exemplo acima os valores “John Smith” para a propriedade “*name*” e “27” para a propriedade “*age*”).

Um *slogan* que se refere a Web é “*Anyone can say Anything about Any topic (AAA)*” (ALLEMBERG; HENDLER, 2011), ou seja, “Qualquer um pode dizer Qualquer coisa sobre Qualquer assunto”. RDF está de acordo com este *slogan* porque permite que qualquer fonte se refira a recursos em quaisquer *namespaces*. Mesmo um único triplo (*subject/predicate/object*) pode se referir a recursos em vários *namespaces*.

Como um modelo de dados, RDF fornece uma clara especificação do que tem que ser feito para mesclar informações de múltiplas fontes. RDF por si só não fornece algoritmos ou tecnologia para implementar esses processos. As tecnologias necessárias são: *RDF Parser/Serializer*; *RDF Store*; *RDF Query Engine*; *Application* (ALLEMBERG, HENDLER, 2011). Neste texto, serão abordados as questões referentes a estas tecnologias. O objetivo aqui é o de fazer uma análise de como esta proposta de organização dos recursos da Web, no contexto da Web de Dados, que tem como base o RDF, está relacionada com as práticas da Representação Descritiva abordada na Ciência da Informação.

⁴ *String* é um conjunto de caracteres.

Uma das aplicações RDF é o SKOS um sistema para a criação de esquemas de conceito.

3 SKOS

Simple Knowledge Organization System (SKOS) é, como o nome representa, um sistema simplificado para a organização do conhecimento. Trata-se de um modelo proposto para expressar esquemas de conceitos. Segundo Isaac e Summers (2009) o modelo serve para expressar vários tipos de esquemas de conceitos tais como: tesouros, sistemas de classificação, listas de cabeçalhos de assunto, taxonomias, terminologias, glossários. Miles e Brickley (2005b) definem esquemas de conceitos como sendo um conjunto de conceitos, que, opcionalmente podem incluir declarações sobre os relacionamentos semânticos entre eles.

Trata-se de uma aplicação em RDF resultante de uma pesquisa do *Semantic Web Advanced Development for Europe Project* que está associado à *Semantic Web Activity* do W3C (W3C SWA).

O modelo é composto por termos num vocabulário denominado *SKOS Core Vocabulary* que é um conjunto de propriedades e classes utilizadas para expressar o conteúdo e estrutura de um esquema de conceitos em RDF.

O RDF é o modelo padrão para intercâmbio de dados na Web recomendado pelo W3C e, portanto, sendo o SKOS escrito em RDF, é possível tornar os esquemas de conceito expressos neste modelo passíveis de serem lidos por agentes inteligentes conforme as recomendações da W3C.

4 APLICAÇÕES EM RDF

Nesta seção apresentam-se as iniciativas ou práticas da Organização da Informação, tanto no que tange a Representação Descritiva quanto a Representação Temática, que já vem sendo realizadas no contexto da Web Semântica.

Em ambas as áreas (descritiva e temática) o modelo RDF aparece como base para a aplicação das práticas da organização da informação conforme segue.

Como destacado anteriormente, a base da Web Semântica é o RDF que, para Manola e Miller (2004) é uma forma de representação dos metadados sobre recursos Web (tais como título, autor, data de modificação de uma página da Web, *copyright*, etc).

Lendo esta definição à luz dos conceitos apresentados relativos à Web Semântica fica clara a relação da práxis das unidades de informação na descrição de itens, que se fundamenta nas teorias da Ciência da Informação, com o que se propõem para a Web de Dados.

Concorda-se com Mey (1995, p. 6), quando afirma que “a riqueza da catalogação repousa nos relacionamentos entre os itens estabelecidos, de forma a criar alternativas de escolha para os usuários”. Enquanto Mey (1995) assevera que a riqueza da catalogação está nos relacionamentos dos itens, a base da Web de Dados está nos Dados Lincados, ou seja, em ambas as áreas de atuação, tanto dos profissionais catalogadores quanto os que propõem a Web Semântica, há a preocupação em dar aos usuários formas mais ricas e amplas de recuperação da informação.

O uso de códigos normalizados na descrição bibliográfica é fundamental por regularizar e padronizar procedimentos em sistemas de informação, especialmente de médio e grande porte, possibilitando maior consistência, eficiência e qualidade ao sistema, ao fornecer registros coerentes entre si; por viabilizar a cooperação catalográfica entre unidades de informação; por diminuir a necessidade de tomar decisões individuais para cada documento a ser catalogado e por facilitar os processos de informatização.

A proposta do W3C é a de que todos os dados e metadados constantes ou citados na Web possam ser processados por máquinas e, para tanto, sugerem como fundamento o RDF que é um modelo de dados para a **descrição** de recursos da Web. Portanto, a base da web de Dados é a “Representação Descritiva” dos recursos a partir de um modelo, em que os metadados são compostos de declarações formadas por simples triplos (*Subject/Predicate/Object*).

Mudanças nos tradicionais códigos de catalogação, modelagem de vocabulários, tesauros, índices, etc, deverão ser formatados dentro dos padrões propostos pelo W3C para a Web Semântica. Inicialmente deve-se pensar em como estão organizados os bancos de dados bibliográficos atualmente. Estes bancos de dados resultam nos catálogos online (OPAC) que são um produto da descrição dos registros bibliográficos. Esses registros bibliográficos são compostos de conjuntos de padrão de metadados padrão tais como o formato MARC (presente nos OPACs) e o Dublin Core (conjunto de metadados adotado pela OAI como a base da interoperabilidade dos repositórios).

Para que os dados dos catálogos bibliográficos possam fazer parte da Web de Dados, é necessário que estes sejam processáveis por máquinas e que tenham RDF como modelo de

descrição. Coyle (2010b, p.15) argumenta que, apesar de o MARC estar estruturado em campos e subcampos que são legíveis por máquinas, tem sua base em linguagem textual, o que dificulta a aplicação das tecnologias da Web Semântica. Porém, o fato de os dados bibliográficos serem fundamentalmente textuais não significa que sejam livres; estes possuem valores controlados, a exemplo do controle de autoridade e de cabeçalhos de assuntos.

Um outro componente importante para que se possa visualizar as contribuições da representação descritiva para a web semântica, são os novos desenvolvimentos da área: o *Functional Requirements for Bibliographic Records*, conhecido como modelo FRBR⁵ (um dos principais objetos de estudo na área da Representação Descritiva) e o *Resource Description and Access* (RDA)⁶.

Supõe-se que para incluir os tradicionais catálogos no ambiente da Web de Dados, ou seja, transformar os dados bibliográficos em dados ligados na Web, faz-se necessário mudar os dados já existentes que se utilizam de MARC ou DC, por exemplo, em dados modelados em RDF. Deve-se, ainda, tornar o modelo conceitual FRBR e o código de catalogação RDA, recentemente publicado como sucessor do AACR, em base RDF. Além do que, escrever os vocabulários controlados existentes em linguagens propostas pelo W3C tais como RDF Schema, SKOS e OWL.

Uma importante iniciativa é a da *Library of Congress* (LC), que disponibiliza uma lista de cabeçalhos em URI (<http://id.loc.gov/>), ou seja, para cada cabeçalho, foi criado um URI, que é um identificador único (base para as declarações em RDF). Exemplo para o nome Jorge Amado: para Amado, Jorge, 1912-2001 a URI é <http://id.loc.gov/authorities/names/n50024126>. Portanto, ao se desenvolverem aplicações de dados bibliográficos em RDF pode-se utilizar o *namespace* dos URIs criados pela LC.

Esta iniciativa vem ao encontro do principal desafio da Web Semântica que são os identificadores únicos para cada recurso. Segundo Coyle (2010b, p.21), para se dar visibilidade aos dados das bibliotecas na Web, primeiramente é necessário que eles tenham identificadores na forma de URIs. Neste sentido, vale destacar as atividades do *Dublin Core Metadata Initiative* (DCMI), organização responsável pela manutenção e desenvolvimento do

⁵ Modelo desenvolvido pela IFLA, resultado de um estudo sobre os requisitos funcionais para registros bibliográficos empreendido de 1992 a 1997. O objetivo do estudo “era produzir uma estrutura (framework) que proporcionasse uma compreensão clara, definida com precisão e comumente compartilhada de como o registro bibliográfico pode proporcionar informação sobre o que se espera que o registro possa arquivar em termos de resposta às necessidades do usuário.” (LE BOEUF, 2007, p.1)

⁶ Proposta do novo padrão que visa fornecer um conjunto de diretrizes e instruções na descrição de recurso e acesso cobrindo todo tipo de conteúdo e mídia.

conjunto de metadados Dublin Core⁷ que criou um identificador para cada um de seus elementos de metadados, disponíveis em <http://purl.org/dc/terms>. Os termos DCMI também estão representados em linguagem RDF Schema (DCMI, 2011).

Relativamente ao RDA, vale destacar a iniciativa do DCMI em conjunto com o *Joint Steering Committee* (JSC). Num encontro promovido, em 2007, pela *American Library Association* (ALA) em conjunto com a *British Library*, os membros do DCMI e do JSC se dispuseram a criar uma expressão do RDA em RDF incluindo as associações deste com as entidades e relacionamentos do FRBR. O resultado é um registro online de RDA em RDF que é a primeira definição de dados bibliográficos no formato para Web Semântica⁸ (COYLE, 2010a). Um artigo foi publicado na *D-Lib Magazine* (2010) relatando o processo de criação do *RDA Vocabularies* (HILMANN et al., 2010).

Na Figura 3 a seguir, pode-se visualizar o detalhamento da propriedade *Author (Work)* apresentado no *Open Metadata Registry*:

Figura 3: Propriedade Autor da Obra no *Open Metadata Registry*

Element Sets: RDA Roles	
Elements: Author (Work)	
Detail	Statements History
Metadata +	
Detail	
Label:	Author (Work)
Name:	authorWork
URI:	http://rdvocab.info/roles/authorWork (RDF)
Description:	A person, family, or corporate body responsible for creating a work that is primarily textual in content, regardless of media type (e.g., printed text, spoken word, electronic text, tactile text) or genre (e.g., poems, novels, screenplays, blogs). Use also for persons, etc., creating a new work by paraphrasing, rewriting, or adapting works by another creator such that the modification has substantially changed the nature and content of the original or changed the medium of expression.
Comment:	
Type:	subproperty
Parent:	Creator
Domain:	http://rdvocab.info/uri/schema/FRBRentitiesRDAWork
Range:	
Status:	New-Proposed
Language:	English
Note:	
List Get RDF	

Fonte: <http://metadataregistry.org/schemaprop/show/id/121.html>

Em relação ao FRBR, existem algumas iniciativas para expressá-lo em RDF. Coyle (2010b) cita: “*Expression of Core FRBR Concepts in RDF*” (ver <http://vocab.org/frbr/core.html>) e o FRBRoo que é parte do “*CIDOC Conceptual Reference Model*” (ver http://cidoc.ics.forth.gr/frbr_inro.html).

⁷ Dublin Core é um conjunto de metadados criado para a descrição de recursos da Web e que é atualmente a base para a interoperabilidade dos repositórios digitais (temáticos ou institucionais).

⁸ *Registry of RDA in RDF* (<http://metadataregistry.org/rdabrowse.htm>)

A primeira iniciativa é um vocabulário que inclui as classes de entidades dos grupos 1, 2 e 3 do FRBR, bem como as propriedades correspondentes para os relacionamentos entre essas entidades. A segunda, o FRBRoo, é uma ontologia criada para a captura e representação da semântica subjacente das informações bibliográficas e para facilitar a integração, mediação e intercâmbio de informações bibliográficas e de museus. Ambas consideram necessário incluir mais algumas entidades no modelo FRBR para poder formatá-lo adequadamente em FRBR.

No estudo realizado pode-se averiguar alguns exemplos nos quais as práticas da Ciência da Informação já tem sido colocadas no contexto da Web Semântica dentro da proposta da Representação Temática. Destacam-se exemplos de tesouros, da CDD e CDU no modelo RDF (SKOS).

Miles e Brickley (2005a) exemplificam com um extrato do *UK Archival Thesaurus* (UKAT), conforme figuras a seguir. Na figura 4 visualiza-se a representação textual do termo e seus relacionamentos:

Figura 4: Extrato do tesaurus UKAT (termo *Economic Cooperation*)

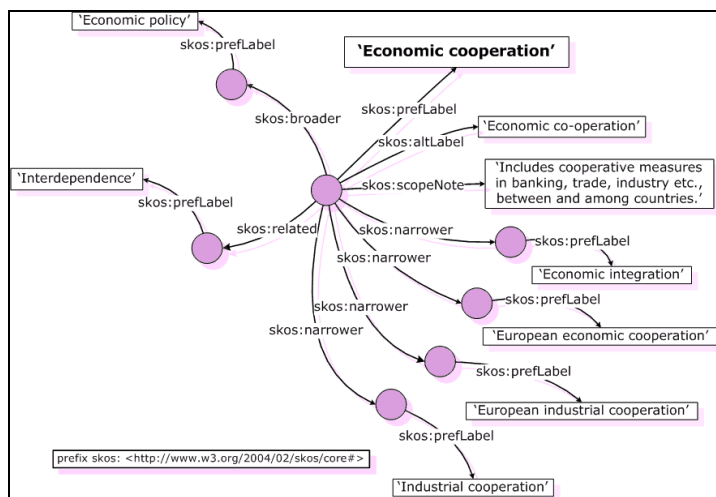
Term:	Economic cooperation
Used For:	Economic co-operation
Broader terms:	Economic policy
Narrower terms:	Economic integration European economic cooperation European industrial cooperation Industrial cooperation
Related terms:	Interdependence
Scope Note:	Includes cooperative measures in banking, trade, industry etc., between and among countries.

Fonte: Miles e Brickley (2005a).

Na figura 5, a representação do termo e seus relacionamentos em gráfico *RDF* (RDF *graph*) usando o *SKOS Core Vocabulary*.

No *RDF graph* são representados os triplos RDF (Subject/Predicate/Object), ou seja, Recurso/Propriedade/Valor (CATARINO; SOUZA, 2012).

Figura 6: RDF *graph*: extrato UK *Archival Thesaurus*



Fonte: Miles e Brickley (2005a).

Neste exemplo o SUBJECT/RECURSO é o termo adotado: ***Economic Cooperation***. Suas propriedades (PREDICATE) são representadas por termos do *SKOS Core Vocabulary*: *Term* (skos:prefLabel), *Used For* (skos:altLabel), *Broader Terms* (skos:broader), *Narrower Terms* (skos:narrower), *Related Terms* (skos:related) e *Scope Note* (skos:scopeNote). Os Valores (OBJECT) são os termos relacionados, por exemplo o termo *Economic Co-operation*; este termo é o valor corresponde à propriedade skos:altLabel (*Used For*).

O exemplo apresentado anteriormente representa uma aplicação do SKOS para um tesouro. No entanto, cabe ressaltar que SKOS pode ser usado para outros tantos esquemas de conceitos. Citam-se aqui outros dois projetos SKOS: a CDD e a CDU nos moldes de dados ligados (*linked data*) da Web Semântica.

O primeiro projeto é uma ação do OCLC (2009). O grupo de trabalho Dewey da OCLC trabalha no sentido de aplicar os princípios de Dados Lincados na Classificação Decimal de Dewey, criando um “Serviço de Terminologia” legível por humanos e/ou por máquinas.

Figura 6: Classe 640 Dewey no Dewey.info

Dewey Decimal Classification	
600	Tecnologia
640	Gerenciamento da casa e família
640	Gerenciamento da casa e família
641	Alimentos e bebidas
642	Refeições e serviços de mesa
643	Casa e equipamentos da casa
644	Serviços da casa
645	Mobiliários da casa
646	Costura, vestuário e vida pessoal
647	Gerenciamento de casas públicas
648	Cuidados da casa
649	Criação de filhos e cuidados domiciliares de pessoas

Fonte: OCLC (2009).

Este serviço disponibiliza um URI para cada conceito da Classificação Decimal de Dewey. Pode-se ver o URI para a classe 640 da CDD (<http://dewey.info/class/64/2009-08/about.pt>). Esta página segue os princípios dos Dados Lincados na medida em que utiliza o sistema SKOS, que por sua vez é em RDF, este último, a base para a Web Semântica.

A seguir uma parte do código fonte da página (representada na figura 7) onde se vê os termos do *SKOS Core Vocabulary* para a descrição dos conceitos. Neste código pode-se ver a representação da classe 640 (Gerenciamento da Casa e Família), relacionando-a com a classe mais geral, 600 (Tecnologia).

Figura 7: Código Fonte: URI classe 640 Dewey

```
<div about="http://dewey.info/class/64/2009/08/about.pt" typeof="skos:Concept">
<span class="edition">http://dewey.info/scheme/e22/</span>
<div rel="skos:broader">
  <span about="http://dewey.info/class/6/2009/08/about.pt" typeof="skos:Concept">
    <a resource="http://dewey.info/class/6/2009/08/about.pt" property="skos:notation"
datatype="ddc:Notation" href="http://dewey.info/class/6/about">600</a>
    <a resource="http://dewey.info/class/6/2009/08/about.pt" property="skos:prefLabel" xml:lang="pt"
href="http://dewey.info/class/6/2009-08/about.pt">Tecnologia</a>
  </span>
</div>
<span class="notation" property="skos:notation" datatype="ddc:Notation">640</span>
<a class="at" resource="http://dewey.info/class/64/2009/08/about.pt" property="skos:prefLabel"
xml:lang="pt" href="http://dewey.info/class/64/2009-08/about.pt">Gerenciamento da casa e família</a>

</span>
```

Fonte: OCLC (2009).

Outro projeto que exemplifica o uso do SKOS para um sistema de classificação decimal é o projeto da *UDC Consortium*, denominado *UDC Summary Linked Data*.

Figura 8: UDC Summary Linked Data: classe 02 Biblioteconomia

UDC Summary Linked Data		Português (Portuguese)
URI	http://udcdata.info/014810	
Notation	02	
Caption	Biblioteconomia	
See also	002 Documentação. Livros. Escritos. Autoria 01 Bibliografia e bibliografias. Catálogos	
Broader class	0 CIÊNCIA E CONHECIMENTO. ORGANIZAÇÃO. INFORMÁTICA. INFORMAÇÃO. DOCUMENTAÇÃO. BIBLIOTECONOMIA. INSTITUIÇÕES. PUBLICAÇÕES	
Narrower classes	021 Função, valor, utilidade, criação, desenvolvimento de bibliotecas 022 Localização da biblioteca, edifícios. Equipamento 023 Administração da biblioteca. Pessoal 024 Relações da biblioteca com o público. Regulamentos sobre o uso da biblioteca 025 Departamentos administrativos das bibliotecas 026 Bibliotecas especializadas 027 Bibliotecas gerais	

Fonte: UDC Consortium (2011).

UDC Summary Linked Data contém 2.400 classes principais (*top level*) da Classificação Decimal Universal (CDU). No site da UDC *Linked Data* é possível visualizar as Tabelas Auxiliares e a Principal. Cada uma das classes é representada num URI específico, conforme exemplo na figura 8, da classe 02 biblioteconomia, <http://udcdata.info/014810>.

No mesmo site ainda é disponibilizado um arquivo para *download* contendo a UDC *Linked Data* em SKOS. A seguir, uma parte deste arquivo contendo apenas a notação 02 Biblioteconomia.

Figura 9: UDC Linked Data, SKOS, 02 biblioteconomia

```
<skos:Concept rdf:about="http://udcdata.info/014810">
<skos:inScheme rdf:resource="http://udcdata.info/udc-schema"/>
<skos:broader rdf:resource="http://udcdata.info/013358"/>
<skos:notation rdf:datatype="http://udcdata.info/UDCnotation">02</skos:notation>
.....
<skos:prefLabel xml:lang="pt">Biblioteconomia</skos:prefLabel>
.....
<skos:related rdf:resource="http://udcdata.info/000002"/>
<skos:related rdf:resource="http://udcdata.info/000001"/>
</skos:Concept>
```

Fonte: UDC Consortium (2011).

5 CONSIDERAÇÕES FINAIS

A Ciência da Informação estuda as questões relativas à captação, tratamento, organização e recuperação da informação, em quaisquer meios e/ou formatos. A partir da aplicação das tecnologias da informação, em especial as redes de computadores, com destaque à Web idealizada na década de 1990, estas questões da Ciência da Informação e da Web se entrelaçam, não sendo hoje possível estudá-las totalmente em separado.

Como pesquisadores da subárea da Organização da Informação e formadores de profissionais que atuarão num mercado de trabalho totalmente voltado para as redes; cabe a nós disseminarmos conhecimentos voltados à prática da organização de recursos da Web.

Verificou-se que já existem ações, tanto na Representação descritiva, quanto na Representação Temática, que visam inserir os dados bibliográficos e sistemas de organização do conhecimento, no contexto da Web de Dados, ou dos Dados Lincados, que tem como base modelos de descrição e de organização, tais como o RDF e o SKOS.

A representação descritiva é uma ação que está presente na Web Semântica na medida em que a base para a sua existência está no uso do RDF (um modelo de descrição). Portanto, a práxis do profissional da informação que cataloga recursos com base nos códigos e modelos que norteiam este fazer (ISBDs, AACR, RDA, FRBR, MARC, etc.), pode ser aplicada para colaborar com o projeto do W3C. A chave está na conversão dos dados bibliográficos atuais em descrições baseadas no modelo RDF.

Os catálogos poderão estar com seus metadados disponíveis por meio dos mecanismos de busca, e este seria o caminho contrário, ou seja, os usuários que estão realizando suas pesquisas nos mecanismos de busca seriam direcionados aos dados lincados das bibliotecas. Transformar os dados bibliográficos em dados lincados não enriqueceria apenas os serviços das bibliotecas ofertados aos seus usuários, mas, inclusive, estará enriquecendo a própria Web, na medida em que será possível, por meio dos mecanismos de busca, acessar os dados bibliográficos presentes nos catálogos de milhares de bibliotecas do mundo todo.

O conceito de Representação Temática é utilizado para se referir a representação dos conceitos de recursos de informação, sendo os vocabulários, produtos resultantes do processo da Representação Temática. Já no contexto da Web Semântica, utiliza-se o conceito de vocabulários de uma forma mais geral: (vocabulários) são utilizados para definir termos (e seus relacionamentos) para descrever e representar uma área do conhecimento, ou para serem adotados numa aplicação específica (W3C, 2012).

Destaca-se o SKOS como um sistema para a organização do conhecimento, que é utilizado para expressar esquemas de conceitos tais como: tesouros, sistemas de classificação, listas de cabeçalhos de assunto, taxonomias, terminologias, glossários e que, opcionalmente, podem incluir declarações sobre os relacionamentos semânticos entre eles. É uma aplicação em RDF resultante de uma pesquisa do *Semantic Web Advanced Development for Europe Project* que está associado à *Semantic Web Activity* do W3C (W3C SWA).

Neste estudo apresentam-se três iniciativas SKOS: um tesauro (UK Archival Thesaurus), a CDD e a CDU. Observa-se, portanto, que já existem iniciativas básicas, tanto na Representação Descritiva quanto na Temática, áreas da Ciência da Informação, para a inserção dos dados bibliográficos e de recursos da Web no contexto da Web de Dados. Isso reforça o questionamento inicial desta pesquisa, de que é imprescindível a formação do Arquivista, do Bibliotecário, enfim, do profissional formado nas escolas da área de Ciência da Informação, nos conteúdos relativos às recomendações do W3C para a Web Semântica.

Como poderá um profissional bibliotecário atuar na organização da informação na Web se não tiver domínio do modelo RDF, do sistema de organização do conhecimento SKOS, na metodologia de construção de Ontologias?

Um profissional formado pela Ciência da Informação deve ter o domínio das práticas da Organização da Informação e saber relacioná-las com o que se tem desenvolvido na Web. Esta é uma situação irreversível pois cada vez mais os conteúdos e dados estarão sendo migrados ou gerados nativamente na Web.

Os resultados desta pesquisa tem sido disseminados em publicações e eventos. O grupo tem, ainda, alguns projetos de disseminação em periódicos, anais e livros. Espera-se que desta forma se possa contribuir para que nossos alunos e profissionais estejam melhor engajados no contexto da Web de Dados, ou seja, vamos contribuir com o projeto da Web Semântica.

REFERÊNCIAS

ALLEMANG, Dean; HENDLER, Jim. **Semantic web for the working ontologist: effective modeling in RDFS and OWL**. 2nd. ed. Amsterdam : Morgan Kaufmann, 2011.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, may 2001. Disponível em: < <http://www.med.nyu.edu/research/pdf/mainim01-1484312.pdf> >. Acesso em: 10 de abr. 2005.

CATARINO, M.E.; SOUZA, T. B. A representação descritiva no contexto da web semântica. **Transinformação**, Campinas, v.24, n.2, maio/ago 2012, p.77-90.

COYLE, Karen. **RDA vocabularies for a Twenty-First-Century Data Environment**. Library Technology Reports, feb./mar., 2010a.

COYLE, Karen . **Understanding the Semantic Web: bibliographic data and metadata**. Library Technology Reports, jan., 2010b.

DCMI (Dublin Core Metadata Initiative). **DCMI Terms declarations represented in RDF schema language**. DCMI, 2011. Disponível em <http://dublincore.org/schemas/rdfs> . Acesso em 18 de outubro 2011.

HILMANN, D. et al. RDA Vocabularies. **D-Lib Magazine**, v.16, n.1/2, jan./feb. 2010. Disponível em: < <http://www.dlib.org/dlib/january10/hillmann/01hillmann.html> >. Acesso em: 20 out. 2011.

INFERÊNCIA. In: MURCHO, Desidério. Dicionário escolar de filosofia.[s.l.] : Plátano Editora, 2011. Disponível em: < <http://www.defnarede.com/i.html> >. Acesso em: 10 out. 2011.

ISAAC, A. ; SUMMERS, Ed. (eds.). **SKOS Simple Knowledge Organization System Primer**:W3C Working Group Note 18 August 2009. Disponível em: < <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/> >. Acesso em 19 nov. 2012.

KING, R. Q&A with Tim Berners-Lee. **BusinessWeek, Ceo guide to Technology**, 9 april 2007. Disponível em: < <http://www.businessweek.com/stories/2007-04-09/q-and-a-with-tim-berners-leebusinessweek-business-news-stock-market-and-financial-advice> >. Acesso em: 14 ago 2007.

LE BOEUF, P. O admirável mundo novo do FRBR. Disponível em: <[http://www.imeicc5.com/download/portuguese/Presentation_s2c_BraveNewFRBRWorld\(PR\)_Port.pdf](http://www.imeicc5.com/download/portuguese/Presentation_s2c_BraveNewFRBRWorld(PR)_Port.pdf)>. Acesso em 9 out. 2011.

MANOLA, F.; MILLER, E. **RDF Primer**. 2004. Disponível em: < <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/> >. Acesso em: 01 de set. 2010.

MEY, E. S. A. **Introdução à catalogação**. Brasília: Briquet de Lemos/Livros, 1995.

MILES, A.; BRICKLEY, D. **SKOS Core Guide**. W3C : 2005a. Disponível em: < <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/> >. Acesso em 20 nov. 2012.

MILES, A.; BRICKLEY, D. **SKOS Core Vocabulary**. W3C : 2005b. Disponível em: < <http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102/pt> >. Acesso em 20 nov. 2012.

OCLC. **Dewey summaries as Linked Data**. OCLC: 2009. Disponível em: <<http://www.oclc.org/dewey/webservices/default.htm> >. Acesso em: 20 nov. 2012.

RDF WORKING GROUP. **Resource Description Framework**. W3C, 2004. Disponível em: <<http://www.w3.org/RDF/> >. Acesso em:10 out. 2011.

UDC Consortium. **UDC Linked Data**. 2011. Disponível em: < <http://udcdata.info>>.

W3C. **Missão do W3C**. W3C, 2012. Disponível em: < <http://www.w3c.br/Sobre/MissaoW3C> >. Acesso em 10 nov. 2012.

W3C. **Semantic Web Terminology**. W3C, 2011. Disponível em:
<http://www.w3.org/2001/sw/wiki/Semantic_Web_terminology>. Acesso em: 10 out 2011.

W3C. **Semantic Web**. W3C, 2010. Disponível em: <
<http://www.w3.org/standards/semanticweb/>>. Acesso em: 01 de Set. 2010.