

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT 8 – Informação e Tecnologia

CONVERGÊNCIA E AMPLIAÇÃO DE DADOS CULTURAIS EM AMBIENTES SEMÂNTICOS

Laís Barbudo Carrasco (Universidade Estadual Paulista - UNESP)

Silvana Aparecida Borsetti Gregorio Vidotti (Universidade Estadual Paulista - UNESP)

José Eduardo Santarém Segundo (Universidade Estadual Paulista - UNESP)

CONVERGENCE AND EXTENSION OF CULTURAL DATA IN SEMANTIC ENVIRONMENTS

Modalidade da Apresentação: Comunicação Oral

Resumo: A pesquisa problematiza o estudo do papel das ontologias na Web Semântica e a maneira que a ontologia CIDOC CRM estrutura, converge e tece a semântica dos dados heterogêneos do Patrimônio Cultural. Dessa forma, a partir do método descritivo, baseado em levantamento bibliográfico e revisão de literatura da área e áreas afins que contemplam o objeto, o trabalho objetiva apresentar a ontologia CIDOC CRM e suas contribuições para a Ciência da Informação no processo de integração, mediação e intercâmbio de informação heterogênea do Patrimônio Cultural em ambientes semânticos. A pesquisa possibilita a compreensão de que a Web Semântica fornece uma estrutura semântica para acesso e compartilhamento de dados e a camada ontologia tece as relações semânticas desta Web de Dados. Além disso, identifica que a ontologia do patrimônio cultural CIDOC CRM pode ser utilizada como referência para o desenvolvimento de projetos nacionais que pretendem fazer com os dados culturais de sistemas heterogêneos do patrimônio cultural sejam integrados e inseridos numa estrutura semântica de dados interligados. Em conclusão, o CIDOC CRM oferece grande contribuição no cenário da Ciência da Informação haja vista que permite a descoberta de recursos entre domínios através de uma arquitetura complexa de conceitos e relacionamentos que pavimenta uma busca ampliada por informação cultural.

Palavras-Chave: Web Semântica; Ontologias; Patrimônio Cultural.

Abstract: The research problematizes the study of the role of ontologies in the Semantic Web and the way that the CIDOC CRM ontology structures, converges and weaves the semantics of the heterogeneous data of Cultural Heritage. Thus, based on the descriptive method, which was based on a bibliographical survey and literature review of the area and related areas that contemplate the object, the objective of the work is to present a CIDOC CRM methodology and its contributions to the Information Science in the integration, mediation and exchange process of heterogeneous information of Cultural Heritage in semantic environments. The research enables an understanding of a Semantic Web, a semantic framework for data access and sharing and an entity layer such as Semantic Data Web relationships. In addition, it is identified as an ontology of cultural heritage. CIDOC CRM can be used as a reference for the development of projects that aim to make the cultural

data of heterogeneous systems of cultural heritage integrated and inserted in a semantic structure of linked data. In conclusion, the CIDOC CRM offers a great contribution within the scenario of Information Science that allows a discovery of resources between domains through a complex architecture of concepts and relationships that pave an expanded search for cultural information.

Keywords: Semantic Web; Ontologies; Cultural Heritage.

1 INTRODUÇÃO

A Web Semântica tem como finalidade o desenvolvimento de um modelo tecnológico que permita o compartilhamento de conhecimento assistido por máquinas. Por conseguinte, a integração das linguagens ou tecnologias *eXtensible Markup Language* (XML), *Resource Description Framework* (RDF), arquiteturas de metadados, ontologias e suas linguagens computacionais como *Web Ontology Language* (OWL), protocolos de interação e recuperação de dados como *Protocol and RDF Query Language* (SPARQL), agentes computacionais, entre outras, favorecerá o aparecimento de serviços Web que garantam a interoperabilidade e cooperação.

No contexto da Web Semântica, a informação é dada com significado explícito, tornando mais fácil para máquinas processarem e integrarem informações disponíveis na rede automaticamente (BERNERS-LEE; HENDLER, 2001).

Isto posto, o que consideramos um problema a ser estudado é o papel das ontologias na Web Semântica e de que maneira a ontologia CIDOC CRM estrutura, converge e tece a semântica dos dados heterogêneos do Patrimônio Cultural no âmbito desta Web de Dados.

O CIDOC (Comitê Internacional para Documentação) CRM (Modelo de Referência Conceitual) é o resultado de mais de dez anos de trabalho do CIDOC Documentation Standards Working Group e do CIDOC CRM SIG que são grupos de trabalho do CIDOC. Desde 2006 é padrão oficial ISO 21127: 2006. Em dezembro de 2014, uma nova versão (com base na versão 5.0.4 do CRM CIDOC) tornou-se disponível: ISO 21127: 2014 (CIDOC CRM, tradução nossa).

Dessa forma, a partir do método descritivo, baseado em levantamento bibliográfico e revisão de literatura da área e áreas afins que contemplam o objeto, este trabalho objetiva apresentar a ontologia CIDOC CRM e suas contribuições para a Ciência da Informação no processo de integração, mediação e intercâmbio de informação heterogênea do Patrimônio Cultural em ambientes semânticos.

2 WEB SEMÂNTICA

Podemos dizer que a Web Semântica é a extensão da Web Atual, isto é, diante da evolução da Web, ou seja, da Web 1.0 para a Web 3.0, houve um desenvolvimento das tecnologias da informação. A Web 1.0 configura-se como um ambiente estático, onde o usuário somente absorve o conteúdo, já na Web 2.0 – a Web Social, o usuário é o protagonista do ambiente, no qual ele participa da interação com a Web, é a Web das redes sociais, das

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

folksonomias, pois o usuário pode ele mesmo inserir seus metadados de conteúdo (SANTOS; ALVES, 2009).

Em contrapartida na Web 3.0, por muitos denominada Web Semântica, é a Web na qual a informação está estruturada com base em tecnologias semânticas, isto é, em tecnologias de interoperabilidade e infraestrutura que permitem que o computador integre e processe informação tendo em vista o seu significado e sua intenção de uso. Sendo assim, o objetivo principal da Web Semântica é desenvolver uma linguagem que os seres humanos possam utilizar para expressar informação que possa ser processada pelos computadores (CHOWDHURY; CHOWDHURY, 2007).

A Web Semântica não trata de uma revolução, mas sim de uma evolução da Web como a conhecemos hoje. Trata-se principalmente da adoção de padrões de metadados e de compartilhamento destes padrões, de forma que possamos melhor utilizar o vasto repositório de informações disponíveis da Web de maneira mais produtiva, ágil e significativa (SOUZA; ALVARENGA, 2004, p.139).

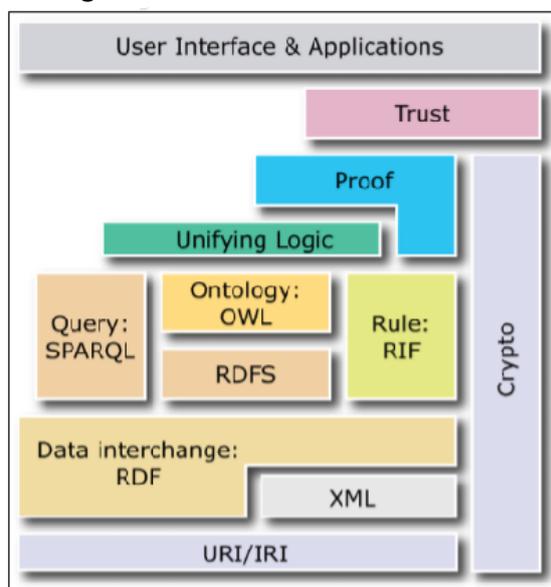
O termo Web Semântica, portanto, representa uma Web de dados ligados/relacionados, isto é, coleções de dados relacionados na Web.

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

A figura 1 apresenta as camadas da Web Semântica que é baseada nas seguintes tecnologias semânticas:

- URI (*Uniform Resource Identifier*) e UNICODE
- XML (*Extensible Markup Language*)
- RDF (*Resource Description Framework*)
- RDF Schema
- Ontologias
- SPARQL.

Figura 1: Camadas da Web Semântica.



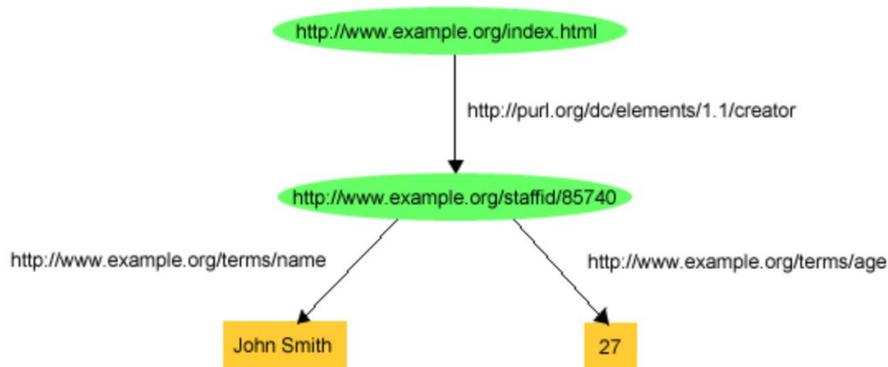
Fonte: W3C (2007).

O URI (*Uniform Resource Identifier*) é um código específico ou identificação designado para recurso da Web que o define unicamente. Em outras palavras, o URI é uma cadeia de caracteres compacta utilizada para identificar ou denominar um recurso na Internet. Um URI pode ser classificado como um localizador (URL – *Uniform Resource Locator*) ou um nome (URN – *Uniform Resource Name*), ou ainda como ambos (BERNERS-LEE, 2005). Em adição, o UNICODE fornece um número único para cada caractere, independentemente da plataforma, do programa e do idioma.

O XML (*Extensible Markup Language*) é uma linguagem que permite a construção de documentos legíveis para seres humanos e que podem ser facilmente tratados por máquinas. Dessa forma, podemos dizer que o XML é um conjunto de regras para a definição de marcadores semânticos, que dividem um documento em partes identificáveis, ou seja, é uma meta-linguagem que define uma sintaxe para ser utilizada na criação de outras linguagens de marcação para um domínio específico, com estrutura e semântica próprias (HAROLD, 1999).

O RDF (*Resource Description Framework*) é um modelo para descrição de recursos da Web, os quais podem ser descritos em *triples*: sujeito, predicado, objeto. É importante ressaltar que o conceito básico do RDF é que tudo seja identificado por meio de URI. Um valor pode ser representado tanto por um URI quanto por uma *string* (literal), conforme demonstrado na figura 2.

Figura 2 – Descrição composta em RDF.



Fonte: Manola e Miller (2004).

O grafo apresentado na figura 2 pode ser descrito de várias maneiras, Manola e Miller (2004), apresentam uma descrição RDF/XML no quadro 1:

Quadro 1 – Descrição 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:extterms="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:Description rdf:about="http://www.example.org/index.html">
    <extterms:creation-date>August 16, 1999</extterms:creation-date>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.example.org/index.html">
    <dc:language>en</dc:language>
  </rdf:Description>
</rdf:RDF>
```

Fonte: Manola e Miller (2004).

Em suma, o RDF é uma estrutura utilizada para descrição e intercâmbio de dados na Web. Sua especificação define um modelo para representação de dados mundiais e sua sintaxe é utilizada para estruturar, representar e intercambiar tal modelo (BALANI, 2005 apud CHOWDHURY; CHOWDHURY, 2007). O RDF apresenta um modo consistente e padronizado para descrever e buscar (*query*) todos os tipos de recursos da Web, desde textos e imagens a arquivos de áudio e vídeos.

De acordo com o W3C (World Wide Web Consortium), no contexto da Web Semântica, o termo busca (*query*) é usado genericamente para se referir às tecnologias e aos protocolos desenvolvidos para a recuperação da informação. O SPARQL (Protocol and RDF Query

Language) é um conjunto de especificações que fornecem linguagens e protocolos para consultar e manipular o conteúdo publicado em RDF na Web (SPARQL, 2013).

No contexto da Web Semântica, as Ontologias são uma estrutura, geralmente um modelo de algum aspecto do mundo, que introduz um vocabulário que descreve diferentes aspectos do domínio que está sendo modelado e fornece uma especificidade explícita do significado pretendido do vocabulário, desse modo, os dados são modelados como um conjunto de relações entre os recursos da Web.

Por fim, note que, segundo Chowdhury e Chowdhury (2007), a Web Semântica não pretende apenas melhorar o acesso aos documentos, ao invés disso, ela é designada para interconectar a gestão de informação pessoal, integração de aplicação corporativa e compartilhamento global de dados comerciais, científicos e culturais. Acesso e compartilhamento de dados é o foco principal da Web Semântica. Por conseguinte, sua tecnologia facilita o acesso à informação baseado na semântica. Assim, pavimentando o caminho para a recuperação semântica da informação.

3 ONTOLOGIAS

O termo ontologia deriva do idioma grego, onto (ser) + logia (estudo), e foi inicialmente difundido dentro dos estudos da Filosofia/Metafísica, para estudar as teorias da natureza da existência. Nesse sentido, lida com a natureza e a organização do ser procurando categorizar o que é essencial e fundamental em determinada entidade. A ontologia faz uma categorização de tudo, conceituando e limitando cada ideia. Por conseguinte, noções de ente, ser, objeto, coisa, existência, essência, realidade, valor, atributo, divindade, necessidade e outras desse tipo são noções ontológicas. Entretanto, sob a óptica da Web Semântica, ontologia pode ser definida como uma teoria lógica sobre uma conceituação específica do mundo real, materializada na forma de um artefato computacional com um compromisso ontológico (MORAIS; AMBRÓSIO, 2007).

Para Gruber (2001), uma ontologia pode ser considerada como uma especificação de uma conceptualização, ou seja, uma descrição de conceitos e relacionamentos que existem entre estes conceitos. Neste caso, uma ontologia é definida como uma especificação formal e explícita de uma conceptualização compartilhada, onde: especificação formal quer dizer algo que é legível para os computadores; explícita são os conceitos, propriedades, relações, funções, restrições e axiomas explicitamente definidos; conceptualização representa um

modelo abstrato de algum fenômeno do mundo real; e compartilhada significa conhecimento consensual (MORAIS; AMBRÓSIO, 2007).

Nesse contexto, para o uso como tecnologia da Web Semântica, entendem-se as ontologias como: artefatos computacionais que descrevem um domínio do conhecimento de forma estruturada, através de: classes, propriedades, relações, restrições, axiomas e instâncias (SANTAREM; CONEGLIAN, 2015).

4 CIDOC CRM

Segundo Gruber (1993, p. 1) “[...] ontologia é uma descrição (como uma especificação formal de um programa) dos conceitos e relações que podem existir formalmente por um agente ou uma comunidade de agentes.” O Modelo de Referência Conceitual (CRM – Conceptual Reference Model) é uma ontologia muito proeminente no contexto dos ambientes digitais de patrimônio culturais.

O CIDOC CRM promove uma compreensão compartilhada de informações do patrimônio cultural, fornecendo um quadro semântico comum e extensível que qualquer informação do patrimônio cultural pode ser mapeada. Destina-se a ser uma linguagem comum para especialistas do domínio e implementadores para formular requisitos para sistemas de informação e servir como um guia para a boa prática de modelagem conceitual. Dessa forma, ele pode fornecer a “cola semântica” necessária para fazer a mediação entre diferentes fontes de informações do patrimônio cultural, como o publicado pelos museus, bibliotecas e arquivos (CIDOC CRM, tradução nossa).

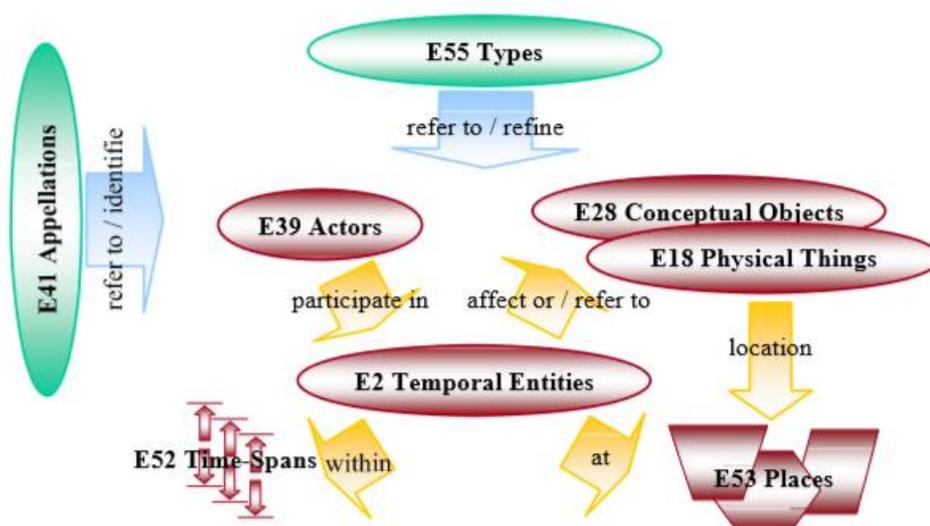
O CIDOC (*Comité International pour la Documentation*) CRM (*Conceptual Reference Model*) surgiu do *CIDOC Documentation Standards Group* no Comitê Internacional de Documentação do Conselho Internacional dos Museus e foi aceita como ISO 21127, em 2006. O objetivo do modelo CRM é fornecer uma linguagem comum para sistemas de informação heterogêneos e permitir a sua integração, apesar de possíveis incompatibilidades semânticas e estruturais. Desse modo, informações do patrimônio cultural podem ser trocadas e recuperadas e as instituições de patrimônio cultural podem tornar seus sistemas de informação interoperáveis, sem comprometer as suas necessidades específicas ou o atual nível de precisão de seus dados (CIDOC CRM).

O CIDOC CRM é uma ontologia de alto nível que permite a integração de informação do patrimônio cultural e sua correlação com informação de museu, biblioteca e arquivo, que, portanto, pode ser facilmente convertida em outros formatos legíveis por máquina, como RDF (*Resource Description Framework*) e XML (*Extensible Markup Language*). Possivelmente a

aplicação mais ambiciosa do CRM é o desenvolvimento de ferramentas de consulta integradas, sistemas de mediação e armazenamento de dados. Atualmente, grande parte das informações armazenadas nos catálogos de biblioteca, inventários de arquivo e coleções de museus encontram-se como sistema de gerenciamento isolado. Diferentes fontes de informação normalmente precisam ser consultadas individualmente e as ligações entre os sistemas são raras. A capacidade de combinar e integrar informações de várias fontes tem o potencial de adicionar um valor significativo aos dados existentes, facilitando a pesquisa e melhorando a qualidade da experiência do usuário (ARTUR; CROFTS; LE BOEUF, 2002).

Há muitos mapeamentos de metadados em XML (*Extensible Markup Language*) para a ontologia CIDOC CRM, uma vez que esta é considerada um dos modelos mais adequados em arquiteturas de integração.

Figura 3 – Conceitos fundamentais do Modelo CIDOC CRM.



Fonte: Doerr, Ore, Stead (2007).

A figura 3 apresenta um esquema de representação de entidades e propriedades do CIDOC CRM, onde entidades espaço temporais conectam coisas, ideias e pessoas através de propriedades, isto é, sentenças verbais. Nesse sentido, os conceitos definidos pelas entidades, os quais são expressos por substantivos, são conectados entre si através das propriedades, as quais são expressas por verbos. Formando, dessa forma, uma sentença formada por sujeito, predicado e objeto. Além disso, as entidades de denominação e de tipo são utilizadas para refinar e detalhar uma entidade principal.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

Quadro 2 – Descrição de Entidades do CIDOC CRM.

Entidades do CIDOC CRM	Descrição
E18 Physical Stuff (Objeto físico)	Grupos de todos os objetos físicos, feitos pelo homem e pela natureza. Por exemplo, riscos, furos, rios e manchas, os quais seriam estranhos se referir apenas como “objetos”.
E28 Conceptual Object (Objeto conceitual)	Grupo de produtos não materiais de nossas mentes e especificamente para permitir o raciocínio sobre a sua identidade, as circunstâncias da sua criação e implicações históricas. Por exemplo, textos, mapas, fotos, música, sons, contos de fadas, sinais, símbolos, padrões, planos, direitos e normas, papel, sinais eletrônicos, marcas, mídia de áudio, pinturas, fotos, memória humana, etc.
E39 Actor (Ator)	Pessoas, indivíduos ou um grupo de pessoas ou organizações, sob o aspecto de seu papel em suas atividades. Por exemplo, o Comitê Central da ISO, o Museu Benaki, em Atenas, Grécia, a Bauhaus em Weimar, Alemanha, Monet, Eu.
E41 Appellation (Denominação)	Todos os nomes em sentido próprio. Códigos ou palavras, sem sentido ou com significado, no roteiro de algum grupo ou codificação de um sistema eletrônico, usados exclusivamente para identificar uma instância específica de alguma categoria dentro de um determinado contexto.
E50 Date (Data)	As datas podem variar em seu grau de precisão. Por exemplo, 1900, abril, 4/04/1959.
E52 Time-Span (Período de tempo)	Determinação de uma gama de datas ou duração sem quaisquer outras conotações, para ser utilizado para confinar períodos, eventos e quaisquer outros fenômenos válidos para um determinado período de tempo. Por exemplo, a partir de 17/12/1993 a 12/08/1996, 14h30 – 16h22, 4 de julho de 1945, 09:30 h, Duração da Dinastia Ming.
E53 Place (Local)	Descrever extensões no espaço, em particular sobre a superfície da terra, no sentido puro da física (independente de fenômenos temporais e matéria).
E55 Type (Tipo)	Representar distinções tipológicas importantes para um determinado grupo de usuários. Por exemplo: peso, comprimento, profundidade são tipos de medição. Retrato, esboço, animação poderia ser tipos de representação. Oral, escrita poderiam ser tipos de linguagem. Excelente, boa, ruim poderiam ser tipos de estado condição.

Fonte: Crofts et al (2015, tradução nossa).

Em suma, o CIDOC CRM contém entidades e grupos lógicos de propriedades. Esses grupos têm a ver com as noções de participação, estrutura, localização, avaliação, identificação, finalidade, motivação, uso e assim por diante. Essas propriedades têm colocado entidades temporais e os eventos em um local central (LIMA, 2008).

4.1 A multiplicidade de associações na orientação ao objeto (*Property Quantifiers*)

Um modelo orientado a objetos é composto por entidades que se relacionam entre si para solucionar um problema proposto. A multiplicidade de associações define quantas instâncias de uma entidade A podem estar relacionadas a uma entidade B por meio dos valores de multiplicidade, apresentados no quadro 3.

Quadro 3: Valores de Multiplicidade.

Valores de multiplicidade	Descrição
<i>many to many</i> (0,n;0,n)	Sem restrições. Uma instância de domínio e uma instância de range dessa propriedade pode ter zero, um ou mais instâncias dessa propriedade.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

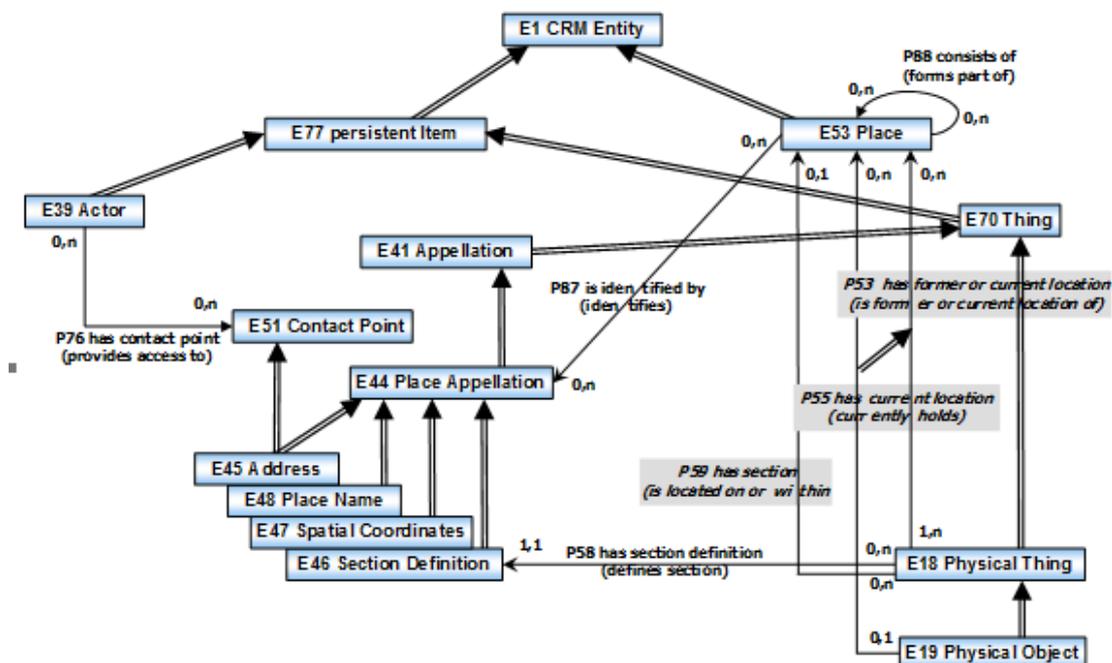
<i>one to many</i> (0,n:0,1)	Uma instância de domínio dessa propriedade pode ter zero, um ou mais instâncias dessa propriedade, mas uma instância de range não pode ser referenciada por mais de uma instância dessa propriedade.
<i>many to one</i> (0,1:0,n)	Uma instância domínio dessa propriedade pode ter zero ou uma instância dessa propriedade, mas uma instância de range pode ser referenciada por zero, uma ou mais instâncias dessa propriedade.
<i>many to many, necessary</i> (1,n:0,n)	Uma instância domínio dessa propriedade pode ter uma ou mais instâncias desta propriedade, mas uma instância de range pode ter zero, um ou mais instâncias desta propriedade.
<i>one to many, necessary</i> (1,n:0,1)	Uma instância domínio dessa propriedade pode ter uma ou mais instâncias desta propriedade, mas uma instância de range não pode ser referenciada por mais de uma instância da propriedade.
<i>many to one, necessary</i> (1,1:0,n)	Uma instância de domínio dessa propriedade deve ter exatamente um exemplo dessa propriedade, mas uma instância de range pode ser referenciada por zero, uma ou mais instâncias desta propriedade.
<i>one to many, dependent</i> (0,n:1,1)	Uma instância de domínio dessa propriedade pode ter zero, um ou mais instâncias da propriedade, mas uma instância de range deve ser referenciada por exatamente uma instância da propriedade.
<i>one to many, necessary, dependent</i> (1,n:1,1)	Uma instância de domínio dessa propriedade pode ter uma ou mais instâncias desta propriedade, mas uma instância de range deve ser referenciada por exatamente uma instância da propriedade.
<i>many to one, necessary, dependent</i> (1,1:1,n)	Uma instância de domínio dessa propriedade deve ter exatamente uma instância dessa propriedade, mas uma instância de range pode ser referenciada por uma ou mais instâncias desta propriedade.
<i>one to one</i> (1,1:1,1)	Uma instância de domínio e uma instância de range dessa propriedade deve ter exatamente uma instância desta propriedade.

Fonte: Crofts et al (2015, tradução nossa).

Diante do exposto, os quantificadores de propriedades são utilizados para a declaração do número permitido de instâncias de uma determinada propriedade que uma instância de range ou domínio pode ter. Essas declarações são ontológicas, ou seja, referem-se à natureza do mundo real descrito e não ao nosso conhecimento atual. Por exemplo, cada coisa tem exatamente um dos pais, mas o conhecimento coletado pode referir-se a nenhum, um ou muitos.

A figura 4 mostra uma vista parcial do CIDOC CRM, a qual representa considerações sobre informações espaciais. Cinco dos principais ramos da hierarquia estão incluídos neste panorama: *E39 Actor*, *E51 Contact Point*, *E41 Appellation*, *E53 Place* e *E70 Thing*.

Figura 4: Considerações sobre informações espaciais.

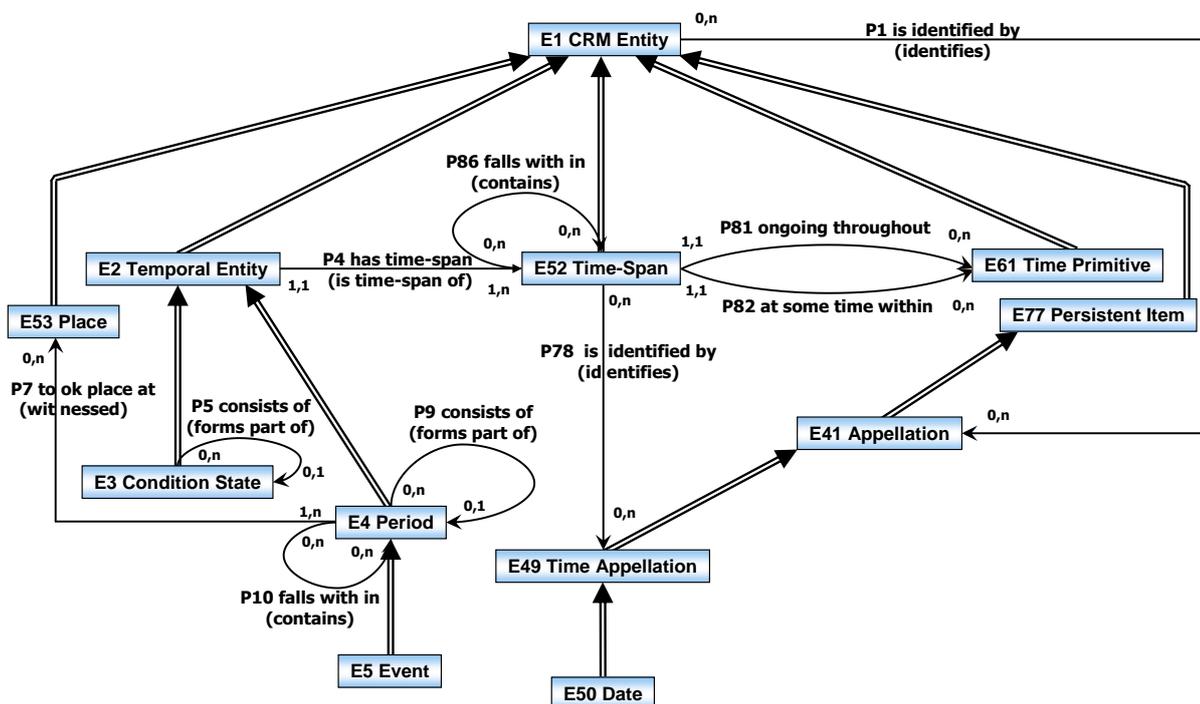


Fonte: Crofts et al (2015).

Como pode ser percebido por meio da figura 4, uma instância do E53 *Place* é identificada por uma instância de E44 *Place Appellation*, que pode ser uma instância de E45 *Address*, E47 *Spatial Coordinates*, E48 *Place Name*, ou E46 *Section Definition*, tais como porão, proa, ou canto inferior esquerdo. Além disso, uma instância do E53 *Place* pode consistir ou fazer parte de uma outra instância do E53 *Place*, permitindo assim uma hierarquia dos 'containers' geométricos a serem construídos. Por fim, uma instância do E45 *Address* pode ser considerada tanto como um E44 *Place Appellation* – uma forma de se referir a um E53 *Place* – e como um E51 *Contact Point* para um E39 *Actor*.

Em adição, a figura 5 mostra como o CIDOC CRM lida com a questão da informação temporal. Em decorrência, quatro dos principais ramos da hierarquia estão incluídos neste panorama: E2 *Temporal Entity*, E52 *Time-Span*, E77 *Persistent Item* e E53 *Place*.

Figura 5: Considerações sobre informações temporais.



Fonte: Crofts et al (2015).

Na figura 5, o E2 Temporal Entity é uma classe abstrata que serve para agrupar todas as classes que contém um elemento temporal, tais como instâncias de E4 *Period*, E5 *Event* e E3 *Condition State*. A propósito uma instância do E52 *Time-Span* é simplesmente um intervalo de tempo que não faz nenhuma referência a contextos culturais ou geográficos, ao contrário do E4 *Period*, que teve lugar em uma instância específica do E53 *Place*. Por fim, note que instâncias do E52 *Time-Span*, às vezes, são identificadas pelas instâncias do E49 *Time Appellation*, muitas vezes sob a forma de E50 *Date*.

4.2 Principais entidades (E) e propriedades (P) do CIDOC CRM

As entidades do CIDOC CRM recebem um nome e um identificador, logo, as entidades recebem um identificador que consiste na letra E seguida por um número. Já as propriedades resultantes também recebem um nome e um identificador, que consiste na letra P seguida de um número. O quadro 4 apresenta as principais entidades do CIDOC CRM.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

Quadro 4: Entidades do CIDOC CRM.

E1	CRM Entity
E2	- Temporal Entity
E4	- - Period
E5	- - - Event
E7	- - - - Activity
E11	- - - - - Modification
E12	- - - - - - Production
E13	- - - - - - Attribute Assignment
E65	- - - - - - Creation
E63	- - - - - - Beginning of Existence
E12	- - - - - - Production
E65	- - - - - - Creation
E64	- - - - - - End of Existence
E77	- Persistent Item
E70	- - Thing
E72	- - - Legal Object
E18	- - - - Physical Thing
E24	- - - - - Physical Man-Made Thing
E90	- - - - - Symbolic Object
E71	- - - - - Man-Made Thing
E24	- - - - - Physical Man-Made Thing
E28	- - - - - Conceptual Object
E89	- - - - - Propositional Object
E30	- - - - - - Right
E73	- - - - - - Information Object
E90	- - - - - - Symbolic Object
E41	- - - - - - Appellation
E73	- - - - - - Information Object
E55	- - - - - - Type
E39	- - Actor
E74	- - - Group
E52	- Time-Span
E53	- Place
E54	- Dimension
E59	Primitive Value
E61	- Time Primitive

Fonte: Crofts et al (2015, tradução nossa).

As entidades do CIDOC CRM são conectadas por meio das propriedades. Note que as propriedades estabelecem uma sentença verbal para os conceitos representados pelas entidades. As principais propriedades do CIDOC CRM estão descritas no quadro 5.

Quadro 5: Propriedades do CIDOC CRM.

Property id	Property Name	Entity – Domain	Entity - Range
P1	is identified by (identifies)	E1 CRM Entity	E41 Appellation
P2	has type (is type of)	E1 CRM Entity	E55 Type
P3	has note	E1 CRM Entity	E62 String
P4	has time-span (is time-span of)	E2 Temporal Entity	E52 Time-Span
P7	took place at (witnessed)	E4 Period	E53 Place
P10	falls within (contains)	E4 Period	E4 Period

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

P12	occurred in the presence of (was present at)	E5 Event	E77 Persistent Item
P11	- had participant (participated in)	E5 Event	E39 Actor
P14	- - carried out by (performed)	E7 Activity	E39 Actor
P16	- used specific object (was used for)	E7 Activity	E70 Thing
P31	- has modified (was modified by)	E11 Modification	E24 Physical Man-Made Thing
P108	- - has produced (was produced by)	E12 Production	E24 Physical Man-Made Thing
P92	- brought into existence (was brought into existence by)	E63 Beginning of Existence	E77 Persistent Item
P108	- - has produced (was produced by)	E12 Production	E24 Physical Man-Made Thing
P94	- - has created (was created by)	E65 Creation	E28 Conceptual Object
P93	- took out of existence (was taken out of existence by)	E64 End of Existence	E77 Persistent Item
P15	was influenced by (influenced)	E7 Activity	E1 CRM Entity
P16	- used specific object (was used for)	E7 Activity	E70 Thing
P20	had specific purpose (was purpose of)	E7 Activity	E5 Event
P43	has dimension (is dimension of)	E70 Thing	E54 Dimension
P46	is composed of (forms part of)	E18 Physical Thing	E18 Physical Thing
P59	has section (is located on or within)	E18 Physical Thing	E53 Place
P67	refers to (is referred to by)	E89 Propositional Object	E1 CRM Entity
P75	possesses (is possessed by)	E39 Actor	E30 Right
P81	ongoing throughout	E52 Time-Span	E61 Time Primitive
P82	at some time within	E52 Time-Span	E61 Time Primitive
P89	falls within (contains)	E53 Place	E53 Place
P104	is subject to (applies to)	E72 Legal Object	E30 Right
P106	is composed of (forms part of)	E90 Symbolic Object	E90 Symbolic Object
P107	has current or former member (is current or former member of)	E74 Group	E39 Actor
P127	has broader term (has narrower term)	E55 Type	E55 Type
P128	carries (is carried by)	E18 Physical Thing	E90 Symbolic Object
P130	shows features of (features are also found on)	E70 Thing	E70 Thing
P140	assigned attribute to (was attributed by)	E13 Attribute Assignment	E1 CRM Entity
P141	assigned (was assigned by)	E13 Attribute Assignment	E1 CRM Entity
P148	has component (is component of)	E89 Propositional Object	E89 Propositional Object

Fonte: Crofts et al (2015, tradução nossa).

Diante do exposto, é importante ressaltar que o CIDOC CRM é um modelo de dados que representa um conjunto de conceitos dentro do domínio do Patrimônio Cultural, no qual os relacionamentos entre os seus principais conceitos são estabelecidos utilizando-se as propriedades descritas no quadro 5.

Diante do exposto, as entidades núcleo do CIDOC CRM descrevem as relações mais fundamentais que conectam coisas, conceitos, pessoas, tempo e local. Dentre as relações,

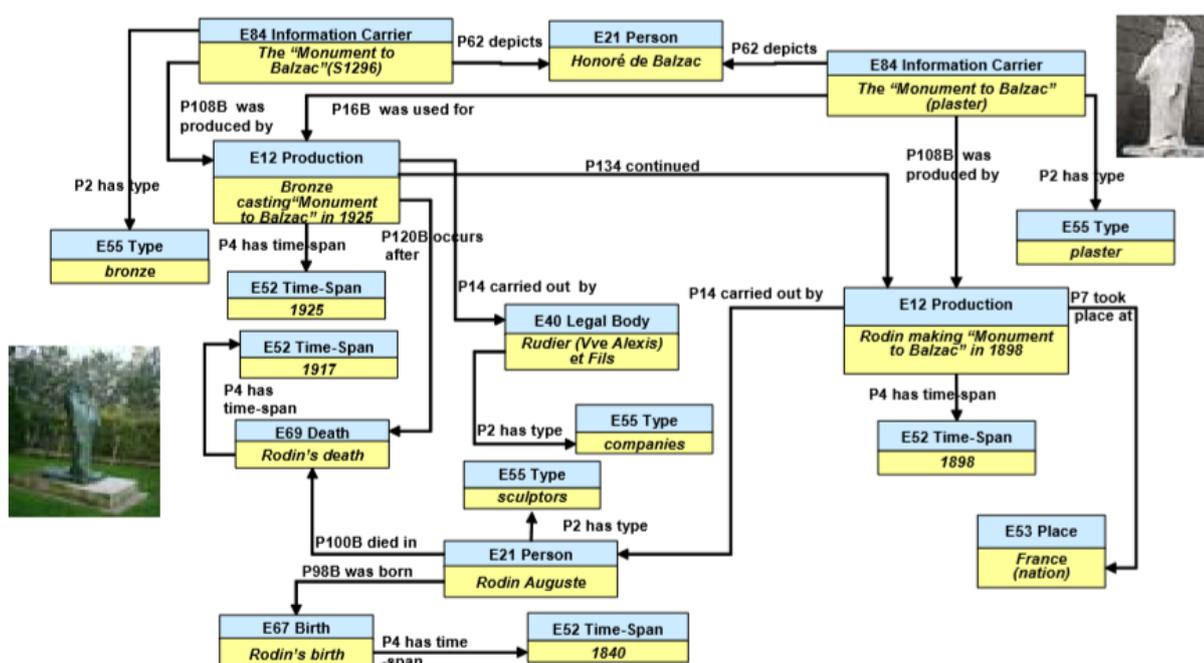
destacam-se identificação, classificação, participação em um evento, hierarquia, referência e semelhança.

4.3 Monumento a Balzac: um exemplo utilizando entidades e propriedades do CIDOC CRM

O escultor francês, Auguste Rodin, passou sete anos se preparando para o Monumento a Balzac por meio de diversos estudos preparatórios que mostram diferentes versões de Balzac. A versão final, em gesso, foi exibida em Paris em 1898 e rejeitada pelos críticos conservadores como um esboço inacabado. Somente anos depois da morte de Rodin, o seu Balzac foi fundido em bronze.

A figura 6 apresenta a representação do Monumento a Balzac, através das entidades e propriedades do CIDOC CRM, em diagrama.

Figura 6 – Monumento a Balzac.



Fonte: Doerr, Ore e Stead (2007).

O quadro 6 apresenta o Monumento a Balzac em descrição XML.

Quadro 6: Monumento a Balzac em XML.

```
<?xml version = "1.0" encoding = "ISO-8859-1"?>
<?xml-stylesheet type="text/xsl" href="crm.xsl"?>

<CRMset>
<CRM_Entity> "Monument to Balzac"

<in_class> E21: Person
<is_identified_by> Rodin Auguste <has_type> Sculptor </has_type></is_identified_by>
```

```
<is_identified_by> Honoré de Balzac <has_note> honored Person</has_note></is_identified_by>
</in_class>

<in_class>E67: Birth
<has_note> Rodin's Birth </has_note>
<was_born> 1840 </was_born>
</in_class>

<in_class>E69: Death
<has_note> Rodin's Death </has_note>
<died_in> 1840 </died_in>
</in_class>

<in_class>E84: Information Carrier
<has_note> Monument to Balzac <has_type> Bronze </has_type></has_note>
<has_note> Monument to Balzac <has_type> Plaster </has_type></has_note>
</in_class>

<in_class>E12: Production
<has_type> Bronze casting "Monument to Balzac" in 1925
<has_note> Production's Year 1925 </has_note></has_type>
<has_type> Rodin making "Monument to Balzac" in 1898
<has_note> Production's Year 1898 </has_note>
<took_place> France </took_place></has_type>
</in_class>

<in_class>E40: Legal Body
<has_note> Rudier (vve Alexis) et Fils <has_type> Companies </has_type></has_note>
</in_class>

</CRM_Entity>
</CRMset>
```

Fonte: Carrasco (2012).

Em síntese, uma produção (evento) de o “Monumento a Balzac” configura o encontro entre o produtor “Rodin” e seu trabalho “Balzac” na “França” em “1898”, que é continuado por outra produção (evento), ou seja, a “Fundição de bronze do Monumento a Balzac” em “1925”, que ocorre após a “morte de Rodin” (evento) em “1917”.

5 CONSIDERAÇÕES FINAIS

As ontologias proporcionam o conhecimento estruturado e uma infraestrutura para possibilitar a integração de bases de conhecimentos, independentes da implementação e constituem uma ferramenta poderosa para suportar a especificação e a implementação de sistemas computacionais de qualquer complexidade (CAMPOS; CAMPOS, 2014).

Sendo assim, as ontologias fazem referência a conceitos e termos que podem ser utilizados para descrever um domínio ou construir uma representação dele. Dessa forma, é importante enfatizar que a ontologia CIDOC CRM, com base em tecnologias semânticas, promove a integração de sistemas heterogêneos digitais do patrimônio cultural.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

Em suma, o papel principal da ontologia CIDOC CRM é permitir o intercâmbio e integração de informações entre fontes heterogêneas de informação do patrimônio cultural. À vista disso, o âmbito de aplicação do CRM pode ser definido como as informações necessárias para o intercâmbio e integração de documentação científica heterogênea de coleções de museus. Além disso, o intercâmbio de informações relevantes com bibliotecas e arquivos, bem como a harmonização do CRM com os seus modelos, se enquadra no âmbito de aplicação do CRM. Soma-se a isso, em um contexto técnico, a ontologia CIDOC CRM pode ser utilizada como uma base para o arquivamento de dados, troca e integração, sendo vista, dessa forma, como um importante contributo para a criação de uma rede global de informação do patrimônio cultural.

A ontologia CIDOC CRM é atuante, principalmente, no cenário internacional das instituições e iniciativas do Patrimônio Cultural, tais como o Museu da Nova Zelândia, *Te Papa Tongarewa*, o qual utilizou o CIDOC CRM para definir as relações entre objetos, pessoas, lugares, categorias e temas narrativos de seu catálogo online e o CLAROS (*Classical Art Research Online Services*), o qual utiliza CIDOC CRM e RDF como base para a combinação de mais de 2.000.000 de registros e imagens realizadas em cinco bases de dados dos principais centros de investigação europeus.

Com esta pesquisa foi possível compreender que a Web Semântica fornece uma estrutura semântica para acesso e compartilhamento de dados e a camada ontologia tece as relações semânticas desta Web de Dados. Além disso, foi possível identificar que a ontologia do patrimônio cultural CIDOC CRM pode ser utilizada como referência para o desenvolvimento de projetos nacionais que pretendem fazer com os que dados culturais de sistemas heterogêneos do patrimônio cultural sejam integrados e inseridos numa estrutura semântica de dados interligados (*Linked Open Data*).

O CRM é utilizado para integrar informação digital em redes de conhecimento de larga escala e significativas que suportam não só o acesso a fontes de documentos, mas também a utilização e reutilização de informações integradas. Em síntese, o CIDOC CRM possibilita a integração semântica de informações de museus, bibliotecas e arquivos, fornecendo, assim, um modelo global e extensível onde os dados provenientes de fontes distintas podem ser mapeados e integrados.

Em conclusão, o CIDOC CRM oferece grande contribuição no cenário da Ciência da Informação haja vista que permite a descoberta de recursos entre domínios através de uma

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

arquitetura complexa de conceitos e relacionamentos que pavimenta uma busca ampliada por informação cultural. As coleções do patrimônio cultural cobrem todo o tipo de material coletado e exibido por museus e instituições relacionadas, tais como bibliotecas e arquivos. Em adição, incluem coleções, sites e monumentos relacionados à história natural, etnografia, arqueologia, monumentos históricos, bem como coleções de artes finas e aplicadas. Portanto, o CIDOC CRM destina-se especificamente a cobrir informações contextuais, ou seja, os antecedentes históricos, geográficos e teóricos onde os itens individuais estão inseridos e que lhes confere grande parte de sua importância e valor.

Como recomendações para o desenrolamento da pesquisa, e como proposta para trabalhos futuros, circunscreve-se a utilidade de a análise da aplicação da norma ISO e ontologia CIDOC CRM nos ambientes digitais do patrimônio cultural disponíveis na Web.

REFERÊNCIAS

ALMEIDA, M; BAX, M. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Revista Ciência da Informação**, 32(3), 2003.

ARTUR, O.; CROFTS, N.; LE BOEUF, P. ELAG presentation Ontologies. Semantic Web and Libraries. **Library Systems Seminar** (26). Rome, 17-19 April 2002. Disponível em: <<http://www.ifnet.it/elag2002/papers/pap9.html>>. Acesso em: 18 abr. 2012.

BERNERS-LEE, T. Semantic Web - XML2000. [2005?]. Disponível em: <<http://www.w3.org/2000/Talks/1206-xml2k-tbl/Overview.html>>. Acesso em: 21 maio 2007.

BERNERS-LEE T.; LASSILA, O.; HENDLER, J. The semantic web. **Scientific American**, New York, v. 5, May 2001.

BREITMAN, K. **Web Semântica: a Internet do futuro**. Rio de Janeiro: LTC, 2005. 190 p.

CAMPOS, L. M.; CAMPOS, M. L. A. Aplicação de dados interligados abertos apoiada por ontologia. In: **ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: além das nuvens, expandindo as fronteiras da Ciência da Informação**, 15., 2014, Belo Horizonte. Anais eletrônicos... Belo Horizonte: UFMG, 2014.

CARRASCO, L. B. **Multilinguality in Heterogenous Digital Cultural Heritage Systems**. 2012. 112 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Letras, Universidade de Coimbra, Coimbra, Portugal, 2012.

CATARINO, M. E.; SOUZA, T. B. A representação descritiva no contexto da web semântica. **Transinformação**, Campinas, Ago. 2012, vol.24, no.2, p.77-90. Disponível em:

<<http://periodicos.puccampinas.edu.br/seer/index.php/transinfo/article/view/766/746>>. Acesso em: 19 nov. 2014.

CHOWDHURY, G. G.; CHOWDHURY, S. **Organizing information from the shelf to the web**. London: Facet Publishing, 2007.

CIDOC CRM - <<http://www.cidoc-crm.org/>>.

CROFTS et al. **Definition of the CIDOC Conceptual Reference Model**. Produced by the ICOM/CIDOC Documentation Standards Group, continued by the CIDOC CRM Special Interest Group. Version 6.0, 2015. Disponível em: <http://www.cidoc-crm.org/docs/cidoc_crm_version_4.2.pdf>. Acesso em: 28 jan. 2015

CROFTS, N. **Museum informatics: the challenge of integration**. Genebra, 2004, 264 p. Tese (Doutorado) – Faculté des sciences économiques et sociales, Universidade de Genebra, Genebra, 2004.

DOERR M., The CIDOC CRM – An Ontological Approach to Semantic Interoperability of Metadata, **AI Magazine**, V. 24, N. 3, pp. 75-92, 2003.

DOERR, M.; ORE, C.; STEAD, S.. The CIDOC Conceptual Reference Model - A New Standard for Knowledge Sharing ER2007 Tutorial. In: 26th **INTERNATIONAL CONFERENCE ON CONCEPTUAL MODELING** (ER 2007), 2007, Auckland. Proceedings. Auckland: CRPIT, 2007. Disponível em: <http://www.cidoc-crm.org/docs/CRM_Tutorial_ER2007.pdf>. Acesso em: 20 maio 2012.

DOERR, M.; STEAD, S. **The CIDOC CRM, a Standard for the Integration of Cultural Information**. Glasgow, Scotland. January 29, 2008. Disponível em: <http://www.uniurb.it/sbc/ist_bal/seminario/carlo%20meghini/CIDOC%20CRM%20in%20production.pdf>. Acesso em: 20 mai. 2012.

GRUBER, T. R. **What is an Ontology?**. Stanford University. 2001. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 27 fev. 2012.

GRUBER, T. R. Towards principles of the design of ontologies used for knowledge sharing. In: **International Journal Human-Computer Studies, Substantial revision of paper presented at the International Workshop on Formal Ontology**, March, Padova, Italy, 1993. Available as Technical Report KSL 93-04, Knowledge Systems Laboratory, Stanford University. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.89.5775>>. Acesso em: 20 jan. 2014.

HAROLD, E. R. **TheXMLBible**. IDG Books, 2 edition, 1999.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

LASSILA, O. **Resource description framework (RDF) model and syntax specification 1.0**. 1999. Disponível em: <<http://www.w3c.org/TR/REC-rdf-syntax>>. Acesso em: 28 abr. 2015.

LIMA, J. A. O., **Modelo Genérico de Relacionamentos na Organização da Informação Legislativa e Jurídica**, PhD, thesis, Brasil, 2008.

MANOLA, F.; MILLER, E. **RDF primer**. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>>. Acesso em: 20 jun 2015.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. **Ontologias: conceitos, usos, tipos, metodologias, ferramentas e linguagens**. Relatório Técnico, Instituto de Informática, Universidade Federal de Goiás, 2007.

PICKLER, M. E. V. Web Semântica: ontologias como ferramentas de representação do conhecimento. **Perspectiva em Ciência da Informação**, Belo Horizonte v.12, n.1, jan./abr. 2007.

SANTAREM SEGUNDO, J. E. ; CONEGLIAN, C. S. . Tecnologias da web semântica aplicadas a organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados. In: José Augusto Chaves Guimarães; Vera Dodebei. (Org.). **Organização do Conhecimento e Diversidade Cultural**. 1ed. Marília: Fundepe, 2015, v. 3, p. 224-233.

SANTOS, P. L. V. A. C.; ALVES, R. C. V. Metadados e Web Semântica para estruturação da Web 2.0 e Web 3.0. **DataGramZero**, Rio de Janeiro, v. 10, n. 6. dez. 2009. Disponível em: <http://www.dgz.org.br/dez09/Art_04.htm>. Acesso em: 21 ago. 2014.

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

SPARQL 1.1 Overview. W3C, 2013. Disponível em <<http://www.w3.org/TR/2013/REC-sparql11-overview-20130321/>>. Acesso em: 28 abr. 2015.

W3C -<<http://www.w3.org/>>

W3C. Layer Cake. 2007. Disponível em: <<https://www.w3.org/2007/03/layerCake.png>>. Acesso em: 1 ago. 2017.