



## Ontologia Cidoc CRM no contexto dos ambientes digitais de patrimônios culturais

*Cidoc CRM ontology in the context of digital cultural heritage environments*

Laís Barbudo Carrasco\*

Manfred Thaller\*\*

Silvana Aparecida Borsetti Gregório Vidotti\*\*\*

### RESUMO

Em decorrência da rápida difusão de informação em ambiente *web*, o acesso multilíngue e a recuperação multilíngue da informação estão se tornando cada vez mais relevantes. Dessa forma, é importante ressaltar que o acesso multilíngue ao conteúdo é usado para aumentar e melhorar as possibilidades dos usuários acessarem os ambientes culturais digitais e, nesse sentido, seu conteúdo poderia ser acessado na língua nativa ou preferencial do usuário. O problema identificado nesta pesquisa é que o patrimônio cultural mundial é geralmente descrito em muitas línguas nacionais diferentes. Multilinguismo em ontologias tem se tornado uma necessidade iminente para as instituições de todo o mundo com valiosos recursos linguísticos. Como a maioria das ontologias é desenvolvida em uma linguagem específica, a obtenção de ontologias multilíngues implica localizá-las ou adaptá-las a uma linguagem concreta e no âmbito de comunidade cultural. Neste artigo, descreveremos a ontologia Cidoc CRM cujo

### ABSTRACT

With the recent rapid diffusion of distributed document bases on the international networks of the web, the matter of multilingual access and information retrieval is becoming increasingly relevant. Thus it is important to emphasize that Multilingual Access to content is used to increase and enhance users' possibilities to access the Cultural Heritage Repositories and their content in their native or preferred language. The issue identified in this research is that the cultural heritage of countries is usually described in many different languages. Multilinguality in ontologies has shown an impending need for institutions worldwide with valuable linguistic resources in different languages. Since most ontologies are developed in one language, obtaining multilingual ontologies implies localizing or adapting them to a concrete language and culture community. In this article we describe the Cidoc CRM ontology, whose goal is to

\* Mestre em EuroMACHS – Patrimônio Europeu, Multimídia e Sociedade da Informação (Universidade de Coimbra – Portugal). Bibliotecária da Força Aérea Brasileira. Endereço: Av. Braz Leme, 3258, Santana, CEP 02022901, São Paulo, SP. Telefone: (11) 22814000. E-mail: laiscarrasco@hotmail.com

\*\* Doutor em História Moderna. Coordenador do mestrado EuroMACHS – Historisch Kulturwissenschaftliche Informationsverarbeitung da Universidade de Colônia. Endereço: Kerpener Str. 30 (Eingang Weyertal, II. Stock, rechts), Köln, Alemanha. Telefone: (49) (221) 4703022. E-mail: manfred.thaller@uni-koeln.de.

\*\*\* Doutora em Educação pela Unesp. Professora do Programa de Pós-Graduação em Ciência da Informação da Unesp. Grupo de pesquisa Novas Tecnologias em Informação. Endereço: Universidade Estadual Paulista Júlio de Mesquita Filho, Departamento de Ciência da Informação. Av. Hygino Muzzi Filho, 737, Campus universitário, CEP 17525900, Marília, SP. Caixa-postal: 180. Telefone: (14) 34021370. E-mail: vidotti@marilia.unesp.br.

objetivo é melhorar a experiência do usuário na busca por conteúdo do patrimônio cultural e, assim, tornar os recursos multilíngues mais operacionais.

**Palavras-chave:** Multilinguismo; Ontologias; Recuperação da Informação; Modelo de Referência Conceitual; Interoperabilidade.

improve the user experience of cultural heritage content by making multilingual features more easily feasible.

**Keywords:** Multilinguality; Ontologies; Information Retrieval; Conceptual Reference Model; Interoperability.

A língua é a base da comunicação entre as pessoas e também faz parte do seu patrimônio cultural. Para muitos, a linguagem tem profundas associações e valores emotivos e culturais enraizados em seu patrimônio literário, histórico, filosófico e educacional. Por esta razão, a linguagem dos usuários não deve ser um obstáculo ao acesso ao patrimônio multicultural disponível no ciberespaço. O desenvolvimento harmonioso da sociedade da informação só é possível, portanto, se a disponibilidade de informação multilíngue e multicultural é incentivada (UNESCO, 2003, p.19, tradução nossa).<sup>1</sup>

## INTRODUÇÃO

Em decorrência da rápida difusão de informação em ambiente *web*, o acesso multilíngue e a recuperação multilíngue da informação estão se tornando cada vez mais relevantes. Dessa forma, é importante ressaltar que o acesso multilíngue ao conteúdo é usado para aumentar e melhorar as possibilidades de os usuários acessarem os repositórios culturais digitais. Nesse sentido, seu conteúdo poderia ser acessado na língua nativa ou preferencial do usuário.

Além disso, pelo fato de os repositórios digitais de patrimônio cultural – como bibliotecas, arquivos, museus e galerias – utilizarem diferentes padrões de metadados para descrever seus recursos de informação, a harmonização de metadados a partir do domínio do patrimônio cultural é um desafio, visto que os modelos de dados são mais projetados nas exigências das comunidades do que em requisitos de interoperabilidade entre as comunidades.

Com o intuito de integrar informações de fontes heterogêneas, ontologias – denominadas tecnologias semânticas – já estão sendo utilizadas, por exemplo, na Europeana.

Europeana é um ambiente digital (biblioteca, museu e arquivo) multilíngue da Europa. Ela oferece aos usuários livre acesso a milhões de livros, pinturas, filmes, sons, objetos de museus e arquivos digitalizados em toda a Europa. A característica mais interessante da Europeana é que ela possui um único ponto de acesso a milhões de materiais que foram digitalizados em toda a Europa, sendo, assim, uma fonte autorizada de informações provenientes de instituições culturais e científicas europeias.

---

<sup>1</sup> “Language is the foundation of communication between people and is also part of their cultural heritage. For many, language has far-reaching emotive and cultural associations and values rooted in their literary, historical, philosophical and educational heritage. For this reason the users’ language should not be an obstacle to accessing the multicultural heritage available in cyberspace. The harmonious development of the information society is therefore only possible if the availability of multilingual and multicultural information is encouraged” (UNESCO, 2003, p.19).

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”. No contexto dos ambientes digitais de patrimônio culturais o Modelo de Referência Conceitual (CRM – Conceptual Reference Model) é uma ontologia muito proeminente.

O Comité Internacional pour la Documentation (Cidoc) propôs uma metodologia eficiente baseada no CRM, ontologia Cidoc CRM, para integrar a riqueza de informação cultural, tornando possível para os usuários a recuperação dos recursos de que necessitam a partir de diferentes coleções. Desse modo, o objetivo do artigo é descrever porque a ontologia Cidoc CRM é a mais adequada para a interoperabilidade entre sistemas heterogêneos do patrimônio cultural.

## **ACESSO E RECUPERAÇÃO MULTILÍNGUE DA INFORMAÇÃO**

Diante do enorme fluxo de informações disponíveis na internet em diferentes idiomas, e levando em consideração a possibilidade de os usuários pesquisarem independentemente de sua língua nativa e também poderem recuperar informação relevante, a questão do acesso e da recuperação multilíngue de informação se tornou de grande importância na atualidade. Assim, o “[...] multilinguismo refere-se à capacidade de uma pessoa utilizar diversas línguas e à coexistência de diferentes comunidades linguísticas em uma área geográfica” (COM, 2005).

Por conseguinte, a World Wide Web deve fornecer sistemas que sejam de fácil navegação, contendo ferramentas flexíveis que ajudem e orientem o usuário na busca por informação. Dessa forma, os usuários podem acessar uma ampla diversidade de fontes de informação sem restrições, que lhes deem a oportunidade de selecionar, descartar e recuperar exatamente aqueles textos que são do seu interesse (CAROL, 1996).

Uma iniciativa importante referente à implementação de recursos multilíngues é a EuropeanaConnect, que “[...] apoia a criação de uma diversa e inclusiva Europeia, facilitando o acesso à cultura por todas as comunidades, indivíduos e representantes de vários grupos culturais e linguísticos”.<sup>2</sup> É importante ressaltar que o portal da Europeia oferece acesso a mais de 20 milhões de objetos digitalizados do patrimônio cultural.

O objetivo da EuropeanaConnect é oferecer componentes essenciais para o desenvolvimento e aprimoramento da Europeia – biblioteca, museu e arquivo da Europa. Dessa forma, conforme o portal da EuropeanaConnect,<sup>3</sup> tal projeto irá:

- fornecer pesquisa e navegação multilíngue.
- enriquecer semanticamente os conteúdos digitais na Europeia;
- desenvolver uma interface espaçotemporal;
- construir uma interface da Europeia para dispositivos móveis;
- estudar o comportamento dos usuários de bibliotecas digitais;

---

<sup>2</sup> Disponível em: < <http://www.europeanaconnect.eu/>. Acesso em: 20 jan. 2014.

<sup>3</sup> Disponível em: < <http://www.europeanaconnect.eu/>. Acesso em: 20 jan. 2014.

- integrar anotação multimídia, informações GIS (Serviços de Informação Geográfica) e de *ebooks on demand*.
- construir e validar um conjunto de licenças para Europeana;
- criar uma infraestrutura de agregação de áudio e adicionar música; e
- implantar componentes fundamentais para a infraestrutura da Europeana.

No mundo digital, há o predomínio da língua inglesa, sendo que os demais idiomas mais falados são: francês, alemão, espanhol e russo. Entretanto, por outro lado, as pessoas preferem pesquisar na World Wide Web em seu próprio idioma, e elas geralmente relatam que sentem falta de informação relevante na sua língua.<sup>4</sup>

Em síntese, o acesso multilíngue permite aos usuários procurarem por informação produzida em diferentes idiomas, sem que seja preciso realizar uma busca em cada um deles.

## A ONTOLOGIA CIDOC CRM

A harmonização de metadados do patrimônio cultural é um desafio porque, geralmente, os modelos de dados são projetados a partir exigências de cada comunidade, nem sempre considerando os requisitos de interoperabilidade entre elas. A interoperabilidade é uma questão importante discutida na contemporaneidade, pois auxilia no gerenciamento de ambientes heterogêneos.

Interoperabilidade é a capacidade de sistemas múltiplos com diferentes plataformas de hardware e software, estruturas de dados e interfaces trocar dados com a mínima perda de conteúdo e funcionalidade. Utilizando esquemas de metadados bem definidos, protocolos de transferência compartilhados e *crosswalks* entre esquemas, os recursos através das conexões podem ser buscados mais facilmente (NISO, 2004, p. 2).<sup>5</sup>

Um modo de fazer os metadados interoperarem seria mapeá-los no âmbito das ontologias. Cidoc Conceptual Reference Model (CIDOC CRM) é uma ontologia do domínio do patrimônio cultural, criada como uma ferramenta para integração da informação.

[Europeana, por exemplo,] fornece um ponto de acesso comum aos objetos do patrimônio cultural digital, em diferentes domínios culturais. Todos esses domínios, a partir de biblioteca, arquivos e museus até os setores audiovisuais, estão usando diferentes formatos de metadados, o que pode ser um obstáculo para a interoperabilidade dentro da Europeana. Práticas precisam ser documentadas em diretrizes e padronizadas, a fim de melhorar a interoperabilidade entre os modelos de dados, e facilitar trabalhos

<sup>4</sup> Disponível em: < <http://www.europeana.eu/portal/>>. Acesso em: 25 mar. 2014.

<sup>5</sup> “Interoperability is the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality. Using defined metadata schemes, shared transfer protocols, and crosswalks between schemes, resources across the network can be searched more seamlessly (NISO, 2004, p. 2).

sobre a qualidade dos dados e enriquecimento (DOERR et al., 2010, tradução nossa).<sup>6</sup>

De acordo com a literatura da área, há muitos mapeamentos de metadados em XML (Extensible Markup Language) para a ontologia Cidoc CRM, uma vez que essa ontologia é considerada um dos modelos mais adequados em arquiteturas de integração.

O Cidoc CRM promove uma compreensão compartilhada de informações do patrimônio cultural, fornecendo um quadro semântico comum e extensível, em que qualquer informação do patrimônio cultural pode ser mapeada. Destina-se a ser uma linguagem comum para especialistas do domínio e implementadores, a fim de 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 as publicadas pelos museus, bibliotecas e arquivos.<sup>7</sup>

Ontologias são compostas de dois conceitos – sujeito e objeto – ligados por um predicado. Tal frase é chamada de *triple* (sujeito – predicado – objeto). Portanto, a estrutura da ontologia é baseada em conceitos estruturados.

Neste artigo, uma ontologia é compreendida como uma descrição formal explícita dos conceitos, em um determinado domínio, propriedades que descrevem as características e atributos de cada conceito, e restrições sobre os atributos. Uma ontologia, juntamente com todas as instâncias individuais de suas classes, constitui uma base de conhecimento. A fronteira entre a noção de ontologia e a de base de conhecimento é um pouco turva (NOY; MCGUINNESS, 2001).

O Cidoc CRM 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. Dessa forma, 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).

---

<sup>6</sup> “[Europeana, for example,] provides a common access point to digital cultural heritage objects across different cultural domains. All these domains, from the library, archives and museum to the audiovisual sectors are using different metadata formats that may be an obstacle to interoperability within Europeana. Practices need to be documented in guidelines, standardized in order to improve the interoperability between the data models and to facilitate works on data quality and enrichment” (DOERR et al., 2010).

<sup>7</sup> “The Cidoc CRM promotes a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modeling. In this way, it can provide the “semantic glue” needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives” (CIDOC CRM).

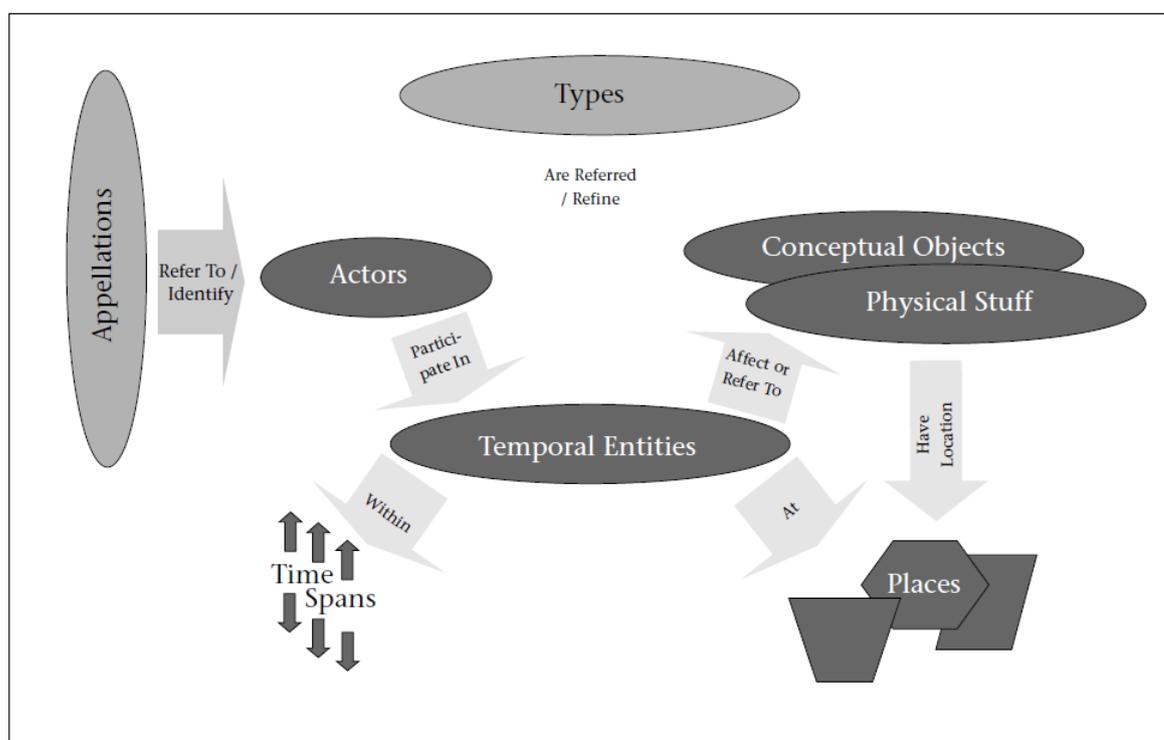
Cidoc CRM é uma ontologia de alto nível que permite a integração de informação do patrimônio cultural e sua correlação com informação de museu, biblioteca e arquivo, que, portanto, pode ser facilmente convertida em outros formatos legíveis por máquina, como RDF (Resource Description Framework) e XML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

Sob a óptica de Crofts (2004, p.107), os objetivos do Cidoc CRM são:

- cobrir todos os aspectos da documentação do patrimônio cultural necessário para troca de informações em um contexto global;
- permitir documentação de conhecimento parcial e contraditório;
- permitir a integração e a troca sem perda semântica entre esquemas relativamente “mais ricos” e “pobres”;
- fornecer uma estrutura extensível e claramente definida para o desenvolvimento futuro.

O âmbito de aplicação do CRM pode ser definido como toda a informação necessária para a documentação científica dos acervos do patrimônio cultural, com vista a permitir um amplo intercâmbio de informações da área e a integração de fontes heterogêneas.

**Figura 1 – Principais entidades do modelo.**



Fonte: Cidoc CRM (DOERR, 2003).

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

**Tabela 1 – Descrição das Entidades do Cidoc CRM.**

Entidades do Cidoc CRM	Descrição
E18 Physical Stuff (Objeto físico)	Grupos de todos os objetos físicos, feitos pelo homem e naturais. 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, usado 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, 04/04/1959.
E52 Time-Span (Período de tempo)	Determinação de uma gama de datas ou duração sem quaisquer outras conotações, a fim de ser utilizada para confinar períodos, eventos, e quaisquer outros fenômenos válidos para um determinado período de tempo. Por exemplos, a partir de 17/12/1993 a 12/08/1996, 14h30 - 16h22 04 julho de 1945, 9h30, Duração da Dinastia Ming.
E53 Place (Local)	Descrever extensões no espaço, em particular sobre a superfície da terra, no sentido puro da física: independentemente de fenômenos temporais e matéria.
E55 Type (Tipo)	Representar distinções tipológicas importantes para um determinado grupo de usuários. Exemplos: peso, comprimento, profundidade, são tipos de medição. Retrato, esboço, animação, poderiam ser tipos de representação. Oral, escrita, poderiam ser tipos de linguagem. Excelente, boa, ruim, poderiam ser tipos de estado ou condição.

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

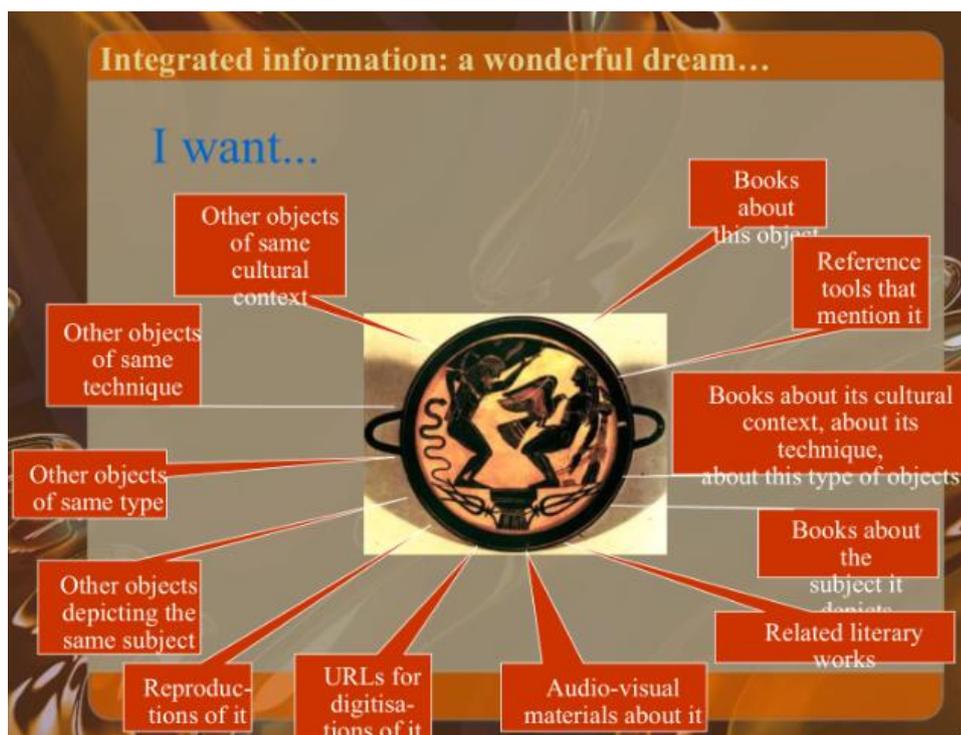
## O MODELO ORIENTADO A OBJETO

Visto que objetos de museu, documentos de arquivo e publicações de biblioteca utilizam diferentes padrões de metadados para a descrição de seus recursos, a ontologia Cidoc CRM, desenvolvida no âmbito da museologia, apresenta uma abordagem orientada para o objeto, ou seja, destaca as relações entre os eventos com o objeto descrito.

Biblioteca, arquivo, museu e galeria são instituições que preservam a memória e organizam o acesso à informação. Por outro lado, de acordo com cada instituição, o papel atribuído ao documento é diferente. Tradicionalmente, as bibliotecas salvaguardam publicações e arquivos de documentos administrativo-legais, e os museus obras-primas (objetos únicos). Assim, cada tipologia documental possui uma maneira específica de ser tratada.

O modelo objeto orientador fornece um mecanismo para esconder detalhes desnecessários. Entidades de baixo nível, por exemplo, que lidam com detalhes específicos, podem ser omitidas, esclarecendo, assim, os aspectos estruturais gerais.

Figura 2 – Relacionamentos com o objeto museológico.

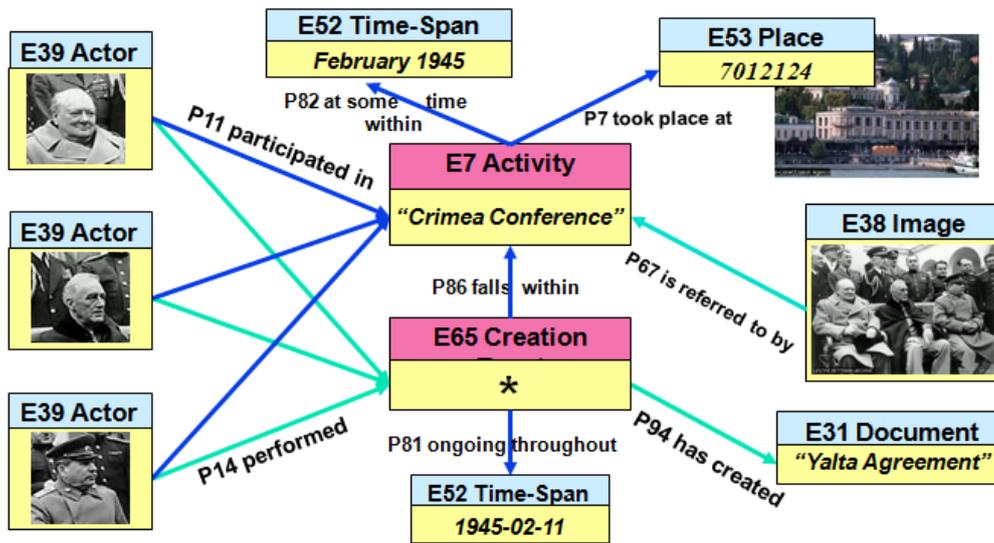


Fonte: Le Boeuf (2006).

A Figura 2 descreve como um objeto museológico pode ser relacionado com diversos recursos. Por exemplo, um artefato pode ser relacionado com outros artefatos do mesmo contexto cultural, com outros objetos da mesma técnica de manufatura, com outros objetos do mesmo tipo, com outros objetos do mesmo assunto, reproduções de tal artefato, endereço eletrônico de digitalizações do mesmo objeto, com materiais audio-visuais sobre ele, livros sobre o objeto, ferramentas de referência que mencionem o objeto, livros sobre o contexto cultural que o objeto foi criado, livros sobre sua técnica, livros sobre objetos do mesmo assunto, livros sobre o assunto do objeto, publicações científicas acerca do objeto, entre outros relacionamentos possíveis.

Em um contexto técnico, a ontologia Cidoc CRM pode ser usada como uma base para o arquivamento de dados, troca e integração, sendo visto, dessa forma como um importante contributo para a criação de uma rede global de informação do patrimônio cultural.

Figura 3 – Conferência de Crimeia.



Fonte: Doerr e Stead (2008).

A Conferência de Crimeia é uma atividade realizada por três atores – Churchill, Roosevelt, Stalin – num local representado pelo código TGN,<sup>8</sup> em um período de tempo específico. Esse episódio é registrado em uma foto (imagem) e relatado em um documento escrito. O seguinte esquema descreve essas informações mediante o uso de entidades e propriedades da ontologia CIDOC CRM, as quais podem ser expressas em XML, da seguinte forma:

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

<CRMset>
<CRM_Entity> Crimea Conference

<in_class> E39: Actor @ pt9 E39: Ator
<is_identified_by> Churchill </is_identified_by>
<is_identified_by> Roosevelt </is_identified_by>
<is_identified_by> Stalin </is_identified_by>
</in_class>

<in_class>E53: Place @ pt E53: Local
<is_identified_by> TGN 7012124
```

<sup>8</sup> TGN (2014).

<sup>9</sup> Os códigos dos idiomas foram retirados da ISO 639 (1989).

```

<has_type> Crimea </has_type>
<has_type> Ukraine </has_type>
</is_identified_by>
</in_class>

<in_class>E7: Activity @ pt E7: Atividade
<is_identified_by> Crimea Conference
<has_type> Conference Region </has_type>
</is_identified_by>

<in_class>E52: Time-Span @ pt E52: Período de tempo
<has_note> February 1945 </has_note>
<has_note> 1945-02-11</has_note>
</in_class>

<in_class> E65: Conceptual creation @pt E65: Criação
conceitual
<took_place_on> 1945-02-11 </took_place_on>
</in_class>

<in_class> E31: Document @ pt E31: Documento
<documents> Photo
<has_note> Actors </has_type>
<has_note> Event </has_type>
</ documents>
<documents> Agreement Text </documents>
</in_class>

<in_class> E38: Image @ pt E38: Imagem
< is_identified_by> Photo </is_identified_by>
</in_class>

</CRM_Entity>
</CRMset>

```

Quando os usuários desejam extrair informações de um sistema, eles precisam fazer uma pergunta, em um padrão de computador. Isso significa que eles têm de formular uma estratégia de busca e, depois, o computador deve ser capaz de encontrar um resultado para a busca. Por outro lado, a maneira mais comum para as pessoas obterem informações é por meio de perguntas em linguagem natural. Assim, as interfaces construídas em linguagem natural facilitam o processo de raciocínio do

usuário e, por serem flexíveis, as pessoas precisam de pouco treinamento para usá-las em interação com um sistema de computador.

A principal utilidade das interfaces em linguagem natural é que elas suportam a vista do domínio pelo usuário e do sistema. Em outras palavras, elas transformam os conceitos do usuário em conceitos efetivamente utilizados pelo sistema. Assim, as interfaces em linguagem natural são sistemas que traduzem uma sentença em linguagem natural para um sistema de busca.

A fim de potencializar a recuperação da informação em sistemas heterogêneos, criamos um mapeamento das entidades do Cidoc CRM (Conceptual Reference Model) com as etiquetas de uma interface de busca. E para atingir um acesso multilíngue, selecionamos entidades do vocabulário da ontologia Cidoc CRM, que utiliza linguagem padrão em língua inglesa, e conectamos com um vocabulário em português. Desse modo, essa correspondência entre etiquetas potencializaria o acesso multilíngue ao conteúdo.

**Tabela 2 – Mapeamento de entidades da ontologia Cidoc CRM em uma interface de busca.**

Interface de Busca		Entidades do Cidoc CRM		Resultado da Busca
Inglês	Português	Inglês	Português	Conferência de Crimeia
Who	Quem	E21 Person E39 Actor	E21 Pessoa E39 Ator	Churchill Roosevelt Stalin
What	O quê	E7 Activity E31 Document	E7 Atividade E31 Documento	Crimea Conference Yalta Agreement
Where	Onde	E53 Place	E53 Lugar	7012124 (TGN)
When	Quando	E52 Time-Span	E52 Período de Tempo	February 1945 1945-02-11

Fonte: Elaborado pelos autores.

Ao traduzirmos a terminologia do vocabulário Cidoc CRM para outros idiomas, criaríamos um ambiente multilíngue para a recuperação da informação. Por exemplo, a entidade E50 Date pode ser traduzida em português como E50 Data, em alemão como E50 Datum, em francês como E50 Date. Já a entidade E35 Title, pode ser traduzida em português como E50 Título, em alemão como E50 Titel, em francês como E50 Titre.

## CONSIDERAÇÕES FINAIS

Multilinguismo em ontologias tem se tornado uma necessidade iminente para as instituições de todo o mundo, com valiosos recursos linguísticos. Como a maioria das ontologias são desenvolvidas em uma linguagem específica, a obtenção de ontologias multilíngues implica localizá-las ou adaptá-las para uma linguagem concreta e comunidade de cultura (GOMEZ; PETERS, 2008).

O multilinguismo tornou-se um importante alvo a atingir porque ele significa muitas possibilidades de linguagens e, a fim de melhorar os resultados da pesquisa, motores de busca associados a uma ontologia multilíngue potencializam o acesso e a recuperação da informação.

É importante notar que o multilinguismo também fornece integração econômica, social, política e cultural dos países. Isso significa que o processo de globalização coloca o desafio do multilinguismo em uma situação de oportunidade.

No Brasil, a ontologia Cidoc CRM ainda não é uma ontologia tão popular quanto na Europa. Como no país ainda há muito que se investir em repositórios digitais do patrimônio cultural, em que o usuário possa localizar facilmente artefatos culturais, como pinturas, gravuras, manuscritos, livros raros, partituras, músicas, esculturas, isto é, um local para interação com a arte, tal ontologia seria a melhor opção para a integração de sistemas heterogêneos do patrimônio cultural.

O vocabulário da ontologia Cidoc CRM pode ser traduzido em muitos idiomas. Portanto, a interface de busca pode utilizar essas etiquetas traduzidas para aprimorar o multilinguismo no sistema. Aqui, é importante mencionar que as etiquetas da interface de busca não precisariam mais ser traduzidas, uma a uma, no *software*, porque a ontologia teria um vocabulário multilíngue. Além disso, usando um vocabulário multilíngue da ontologia Cidoc CRM, a interface de pesquisa se apropriaria das etiquetas traduzidas para o motor de busca e poderia usá-las em vez de traduzir etiqueta por etiqueta no *software* do sistema.

O Cidoc CRM promove acesso multilíngue à informação, bem como proporciona uma recuperação de informação mais elevada para a consulta do usuário. Como a língua não é neutra, o acesso e recuperação multilíngue de informação é um tema para muita reflexão.

Artigo recebido em 18/02/2015 e aprovado em 12/03/2015

## REFERÊNCIAS

ARTUR, O.; CROFTS, N.; LE BOEUF, P. Elag presentation ontologies. Semantic web and libraries. In: LIBRARY SYSTEMS SEMINAR, 26., 2002, Rome. *Proceedings...* Rome: European Library Automation Group, 2002. Disponível em: <<http://www.ifnet.it/elag2002/papers/pap9.html>>. Acesso em: 18 abr. 2012.

BREITMAN, K. *Web semântica: a internet do futuro*. Rio de Janeiro: LTC, 2005.

CIDOC CRM. Disponível em: <<http://www.cidoc-crm.org/>>. Acesso em: 20 fev. 2014.

COMMISSION OF THE EUROPEAN COMMUNITIES. *A new framework strategy for multilingualism*. Brussels, 2006. Disponível em: <<http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52005DC0596>>. Acesso em: 20 fev. 2012.

CROFTS, N. *Museum informatics: the challenge of integration*. 2004. Tese (Doutorado em Ciências Econômicas e Sociais) – Faculté des Sciences Économiques et sociales, Universidade de Genebra, Genebra, 2004.

CROFTS et al. (Ed.). *Definition of the Cidoc Conceptual Reference Model*. Cidoc, Icom, 2005. Disponível em: <[http://www.cidoc-crm.org/docs/cidoc\\_crm\\_version\\_4.2.pdf](http://www.cidoc-crm.org/docs/cidoc_crm_version_4.2.pdf)>. Acesso em: 28 jan. 2015.

DOERR M., *The Cidoc CRM: an ontological approach to semantic interoperability of metadata*. *AI Magazine*, v. 24, n. 3 p. 75-92, 2003.

DOERR, M.; STEAD, S. *The Cidoc CRM: standard for the integration of cultural information*. Glasgow, Jan. 29, 2008. Disponível em: <[http://www.uniurb.it/sbc/ist\\_bal/seminario/carlo%20meghini/CIDOC%20CRM%20introduction.pdf](http://www.uniurb.it/sbc/ist_bal/seminario/carlo%20meghini/CIDOC%20CRM%20introduction.pdf)>. Acesso em: 20 maio 2012.

DOERR et al. *The Europeana Data Model (EDM)*. In: IFLA GENERAL CONFERENCE AND ASSEMBLY, 76., 2010, Gothenburg. *Proceedings...* Gothenburg: Ifla, 2010. Disponível em: <<http://conference.ifla.org/past-wlic/2010/149-doerr-en.pdf>>. Acesso em: 18 fev. 2012.

EUROPEANA. Disponível em: <<http://www.europeana.eu/portal/>>. Acesso em: 18 fev. 2012.

EUROPEANACONNECT. Disponível em: <<http://www.europeanconnect.eu/>>. Acesso em: 18 fev. 2012.

GOMEZ, A.; PETERS, W. *Modelling multilinguality in ontologies*. In: COLING, 2010, Manchester. *Proceedings...* Manchester: Coling, 2008, 67–70 p. Disponível em: <<http://www.aclweb.org/anthology-new/C/Co8/Co8-2017.pdf>>. Acesso em: 30 maio 2012.

GRUBER, T. R. *Towards principles of the design of ontologies used for knowledge sharing*. *International Journal of Human-Computer Studies*, v. 43, n. 5-6, p. 907-928, nov.-dez. 1995. Substantial revision of paper presented at the International Workshop on Formal Ontology, Padova, 1993. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.89.5775>>. Acesso em: 20 jan. 2014.

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

ISO 639 – Code for the Representation of the Names of Languages, 1989. Disponível em: <<http://xml.coverpages.org/iso639a.html>>. Acesso em: 26 jan. 2014

LE BOEUF, P., *Using an ontology-driven system to integrate museum information and library information*. In: SYMPOSIUM ON DIGITAL SEMANTIC CONTENT ACROSS CULTURES, 2006, Paris. *Proceedings...* Paris: Louvre, 2006.

LIMA, J. A. O. *Modelo genérico de relacionamentos na organização da informação legislativa e jurídica*. 2008. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, DF, 2008.

NISO [National Information Standards Organization]. *Understanding metadata*. Bethesda, MD: Niso Press, 2004. Disponível em: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>. Acesso em: 20 fev. 2012.

NOY, N.; MCGUINNESS, D. *Ontology development 101: a guide to creating your first ontology*. Stanford: Stanford University, 2001.

PETERS, C. Guaranteeing multilinguality in the information society. In: THE INFORMATION SOCIETY IN THE EURO-MEDITERRANEAN CONTEXT CONFERENCE, 1996, Antipolis. *Proceedings...* Antipolis: Ercim, 1996. Disponível em: <<http://www.ercim.eu/medconf/papers/peters.html>>. Acesso em: 29 fev. 2012.

TGN. *Thesaurus of geographic names*. Disponível em: <<http://www.getty.edu/research/tools/vocabularies/tgn/>>. Acesso em: 24 jan. 2014.

UNESCO. *Cultural and linguistic diversity in the information society*. Paris: Unesco, 2003. Disponível em: <<http://unesdoc.unesco.org/images/0013/001329/132965e.pdf>>. Acesso em: 15 ago. 2012.