



Encontros Bibli: revista eletrônica de
biblioteconomia e ciência da informação

E-ISSN: 1518-2924

bibli@ced.ufsc.br

Universidade Federal de Santa Catarina
Brasil

Saraiva CONEGLIAN, Caio; Santarem SEGUNDO, José Eduardo
Europeana no Linked Open Data: conceitos de Web Semântica na dimensão aplicada
das Humanidades Digitais
Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, vol. 22,
núm. 48, enero-abril, 2017, pp. 88-99
Universidade Federal de Santa Catarina
Florianópolis, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=14748878008>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica
Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

ARTIGO

Recebido em:
09/07/2016

Aceito em:
10/11/2016

Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 22, n.48, p. 88-99, jan./abr., 2017. ISSN 1518-2924. DOI: 10.5007/1518-2924.2017v22n48p88

Europeana no Linked Open Data: conceitos de Web Semântica na dimensão aplicada das Humanidades Digitais

Europeana in Linked Open Data: Semantic Web concepts in applied dimension of the Digital Humanities

Caio Saraiva CONEGLIAN (caio.coneglian@gmail.com)* - José Eduardo Santarem
SEGUNDO (santarem@usp.br)**

* Mestrando em Ciência da Informação pela Universidade Estadual Paulista "Júlio de Mesquita Filho"

** Professor da Universidade de São Paulo (Ribeirão Preto); Professor do Programa de Pós-Graduação em Ciência da Informação pela Universidade Estadual Paulista "Júlio de Mesquita Filho"

Resumo

O surgimento de novas tecnologias, tem introduzido meios para a divulgação e a disponibilização das informações mais eficientemente. Uma iniciativa, chamada de Europeana, vem promovendo esta adaptação dos objetos informacionais dentro da Web, e mais especificamente no Linked Data. Desta forma, o presente estudo tem como objetivo apresentar uma discussão acerca da relação entre as Humanidades Digitais e o Linked Open Data, na figura da Europeana. Para tal, utilizamos uma metodologia exploratória e que busca explorar as questões relacionadas ao modelo de dados da Europeana, EDM, por meio do SPARQL. Como resultados, compreendemos as características do EDM, pela utilização do SPARQL. Identificamos, ainda, a importância que o conceito de Humanidades Digitais possui dentro do contexto da Europeana.

Palavras-chave: Web semântica. Linked open data. Humanidades digitais. Europeana; EDM.

Abstract

The emergence of new technologies, has introduced means for the dissemination and availability of information more efficiently. An initiative, called Europeana, has been promoting this adaptation of information objects in the Web, and more specifically in Linked Data. Thus, this study aims to present a discussion on the relationship between the Digital Humanities and the Linked Open Data within the context of the Europeana. To do this, we use an exploratory methodology and seeks to explore issues related to the Europeana Data Model, EDM, through SPARQL. As a result, we understand the characteristics of the EDM by using SPARQL. We identified also the importance that the concept of Digital Humanities has within the context Europeana.

Keywords: Semantic web. Linked open data. Digital humanities. Europeana. EDM.



v. 22, n. 48, 2017.
p. 88-99
ISSN 1518-2924



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).

1 INTRODUÇÃO

As iniciativas de Web Semântica estão se tornando mais sólidas e materializadas com o passar dos anos. Destacam-se diversos estudos utilizando os conceitos e tecnologias da Web Semântica no meio acadêmico, em mecanismos de busca, e possivelmente a iniciativa mais sólida no emprego dos recursos da Web Semântica com efetividade são os *datasets* publicados por meio do Linked Open Data (LOD).

Nesse contexto, o LOD pode ser definido como um conjunto das melhores práticas para realizar a publicação e conexão de dados estruturados na Web, para assim, permitir que sejam criadas ligações entre dados de distintas fontes espalhadas em âmbito global (HEATH, 2012).

O projeto contém mais de 570 bases de dados estruturadas conectadas e apresenta um crescimento considerável ao passar dos anos (CYGANIAK; JENTZSC, 2014). Uma dessas iniciativas, que apresentam grande destaque e importância, é a Europeia. A Europeia é um projeto de uma biblioteca virtual de toda Europa, que reúne e integra mais de 50 milhões de objetos culturais, de bibliotecas, de museus e de arquivos. O projeto da Europeia de LOD contém um ambiente de SPARQL *Endpoint*, que possibilita a recuperação dos dados estruturados através de uma linguagem de chamada de SPARQL Protocol and RDF Query Language (SPARQL). Vale ressaltar que, o desenvolvimento da Europeia está inserido no contexto das Humanidades Digitais.

As Humanidades Digitais, apresenta-se como a disciplina capaz de realizar intersecção entre as tecnologias e as ciências humanas, na busca de tornar mais efetivo o acesso e recuperação das informações geradas pelas ciências humanas, por meio da aplicação das tecnologias (SANTAREM SEGUNDO, 2015).

O EDM (*Europeana Data Model*) é modelo conceitual de descrição de recursos utilizado pela Europeia, inclusive na disponibilização de dados no LOD, e se torna pauta de discussão nesta pesquisa no que concerne a temática de Humanidades Digitais e Organização do Conhecimento. Portanto este estudo concentra-se e tem como objetivo principal a análise e identificação do modelo de descrição da Europeia no contexto do LOD, pois trata-se de um dos principais modelos de publicação de dados em grande volume desta natureza. Busca-se ainda, por meio do uso da linguagem SPARQL identificar as principais características do modelo EDM e como os dados da Europeia estão vinculados a outras bases de dados.

A metodologia utilizada está compreendida em um estudo exploratório e aplicado, com natureza qualitativa. Para tal, a pesquisa se dividiu em três momentos: inicialmente explorou literaturas que indicassem o funcionamento do modelo de dados utilizado pela Europeia; na sequência, buscou-se por meio do uso do SPARQL explorar o modelo de dados EDM e; por fim, realizou-se uma discussão, na tentativa de apontar a proximidade que o EDM tem com as Humanidades Digitais, Organização do Conhecimento e Linked Open Data.

2 WEB SEMÂNTICA E LINKED OPEN DATA

A estruturação dos dados na Web é um desafio que vem sendo discutido desde a sua fundação, no começo da década de 1990. Isto porque a Web nasceu descentralizada, como um grande depósito de documentos, dificultando a recuperação das informações, além de tornar impossível que agentes automatizados fossem capazes de compreender o sentido que os dados espalhados pela rede possuíam.

A primeira iniciativa real que visava corrigir esta questão, foi a Web Semântica, que foi proposta em 2001 por Tim Berners-Lee, James Hendler e Ora Lassila. Neste documento, os autores refletem sobre como a Web deve se organizar e se estruturar para que os computadores pudessem compreender o contexto dos dados, e assim, tais máquinas auxiliarem as pessoas em suas tarefas cotidianas. Este texto por mais que apresenta um caráter teórico, começa a inserir algumas tecnologias que serão a base para a estruturação da Web Semântica. A partir disto, a Web Semântica iniciou um processo de maturação, com o desenvolvimento de tecnologias que possibilitassem a implementação desta iniciativa. (BERNERS-LEE; HENDLER; LASSILA, 2001).

Algumas destas tecnologias desenvolvidas, tornaram-se fundamentais dentro da Web para a descrição das informações, como *Uniform Resource Identifier* (URI) e *Resource Description Framework* (RDF). Neste contexto, “Uma Uniform Resource Identifier (URI)

fornece um meio simples e extensível para a identificação de um recurso.” (BERNERS-LEE, et al., 2005, tradução nossa). Uma URI apresenta importância devido a necessidade da identificação fácil dos recursos da Web. Outra característica relevante é a uniformidade dos recursos, que promove a diferenciação dos mesmos, facilitando a compreensão do contexto que um recurso possui.

Outra tecnologia, o RDF, pode ser considerado como a linguagem de descrição da Web. Ferreira e Santos apontam que o modelo RDF “[...] pode ser utilizado para representar recursos de maneira simples, flexível e interpretável, podendo conter representações abrangentes, do ponto de vista cognitivo, e sintéticas para a tomada de decisão dos usuários sobre os recursos buscados.” (FERREIRA; SANTOS, 2013, p. 21). A constatação dos autores indica que o RDF se mostra como fundamental dentro da Web Semântica, por este modelo abranger a necessidade de uma descrição que explicita o contexto dos dados, além de se relacionar às questões relativas a tomada de decisões.

Um outro conceito que apresenta distinta importância para a Web Semântica é o uso de ontologias para a contextualização das informações. Desta forma, “Para o uso como tecnologia da Web Semântica, entende-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 SEGUNDO; CONEGLIAN, 2015, p. 227). Neste sentido, a *Web Ontology Language* (OWL) como linguagem recomendada para a construção de ontologias possibilita a implementação das questões destacadas pelos autores em uma linguagem computacional.

A utilização de RDF e OWL geram dados estruturados que necessitam de linguagens e técnicas específicas para realizar a recuperação dos dados. Assim, o SPARQL, que tem como estrutura de suas consultas o RDF, surge como a principal tecnologia para localizar dados estruturados nas tecnologias da Web Semântica. Berners-Lee afirma que “[...] tentar usar a Web Semântica sem SPARQL é como tentar usar um banco de dados relacional sem SQL. SPARQL torna possível consultar informações de bancos de dados e outras fontes diversas em estado natural, em toda a Web” (W3C, 2008, tradução nossa).

Com o passar dos anos e o aprimoramento das tecnologias que estruturam a Web Semântica, surgiram projetos que tinham a proposta de tornar real ou materializar a Web Semântica. Um destes projetos é o LOD, que tem sido uma das propostas com mais visibilidade no contexto da Web Semântica.

O LOD foi proposto em 2006 por Berners-Lee (2006), que relata que a Web Semântica tem como base a existência de ligações entre os dados, que permite que homens e agentes computacionais sejam capazes de explorar a chamada Web de dados. O autor complementa dizendo que a existência dos dados ligados promove a descoberta dos dados pelos usuários.

Na proposta, Berners-Lee insere quatro princípios do LOD:

1. Usar URIs como nomes para as coisas;
 2. Use HTTP URIs para que as pessoas possam procurar esses nomes;
 3. Quando alguém procura uma URI, fornecer informações úteis, usando padrões (RDF *, SPARQL);
 4. Incluir links para outras URIs para que eles possam descobrir mais coisas.
- (BERNERS-LEE, 2006, tradução nossa).

Estes princípios se baseiam, principalmente, na utilização de URI e de modelos para a representação de dados, como o RDF. A ideia do LOD se encontra em possuir ligações entre os dados, que possuam uma contextualização de forma padronizada. Além disso, os *datasets* do LOD costumam disponibilizar ambientes chamados de SPARQL *Endpoint*, que permitem que os usuários consultem seus dados por meio de consultas de SPARQL.

Um dos projetos de maior destaque no LOD, é a Europeana. Na próxima seção será tratado com maiores detalhes sobre esta iniciativa.

3 EUROPEANA DATA MODEL (EDM)

A Europeana nasceu da necessidade de unir diversos documentos espalhados pelas bibliotecas, museus, arquivos e outros centros culturais espalhados pela Europa, em uma tentativa de disponibilizar em um ambiente digital o acesso a essa grande massa informacional.

Considerando a importância cultural que a Europa possui, devido as interações que tal continente apresenta com o redor do mundo desde os primórdios da civilização, a preservação de todo o conhecimento registrado surgiu como um dos grandes propósitos e motivações para a criação da Europeana (WINER; ROCHA, 2013).

Desta forma, Winer e Rocha (2013) relatam que a partir de 2010, o projeto começou a se tornar realidade, com a criação de um comitê denominado de “Comitê dos Sábios”. Este comitê tinha como intenção fazer as recomendações à Comissão Europeia e instituições culturais europeias, de como disponibilizar e preservar na Internet o patrimônio cultural desse continente, para as próximas gerações. Tais recomendações perpassam por questões de acesso aos documentos, o fomento a iniciativa, tornar a Europeana uma referência ao acesso online ao patrimônio cultural, sustentabilidade, entre outros.

A partir dessas recomendações, o acervo da Europeana atingiu em novembro de 2010 mais de 25 milhões de documentos. Destaca-se que a partir disso a quantidade de documentos vem crescendo continuamente, atingindo em fevereiro de 2016, mais de 49 milhões. O acervo contém basicamente, obras de arte, artefatos, livros, vídeos e sons, apresentando diversas coleções e exposições, que promove um ambiente de bastante interatividade, que promove ao usuário uma ampla quantidade de opções de navegação.

O caráter inovador desse projeto se mostra nos diversos componentes que formam essa biblioteca virtual com destaque para o EDM que utiliza diversas tecnologias da Web Semântica, como URI e RDF.

O Modelo de Dados EDM utiliza diversos vocabulários para formar um modelo que consiga expressar as ligações existentes, entre autores, obras, organizações, direitos autorais, além de outros tipos de informações, contidas em um objeto cultural.

Segundo Europeana (2014, p. 8, tradução nossa),

O Modelo de Dados Europeana (EDM) é destinado a ser um meio de integração para coletar, conectar e enriquecer as descrições fornecidas por provedores de conteúdo da Europeana. Como tal, pode-se dizer que inclui qualquer elemento (como classe ou propriedade) encontrados na descrição de um fornecedor de conteúdos. Dar conta de todos esses elementos é claramente uma tarefa impossível, uma vez que formam um conjunto aberto, ou seja, um conjunto que pode ser estendido como novos fornecedores de juntar-se ao espaço de informação da Europeana.

Os elementos reusados pelo EDM fazem parte dos seguintes vocabulários: RDF e o RDF Schema (RDFS), *OAI Object Reuse e Exchange* (ORE), *Simple Knowledge Organization System* (SKOS), Dublin Core, *W3C Data Catalog Vocabulary* (DCAT) e *Creative Commons* (CC). (EUROPEANA, 2014).

O uso e integração de vocabulários internacionalmente reconhecidos e nascidos de padrões de metadados já sedimentados, torna o EDM um rico modelo de descrição dos objetos da Europeana, fortalecendo a estrutura de descrição semântica dos recursos.

A utilização deste modelo é fundamental na publicação do conjunto de dados da Europeana no LOD, em seu *dataset* “data.europeana.eu”.

A publicação dos dados no LOD, com estrutura organizada e uniforme, é possível pelo modelo de conversão de dados adotado pela equipa da Europeana na transformação dos modelos de descrição encontrados em suas fontes primárias (museus, bibliotecas e arquivos) para inserção dos dados em sua base de dados e posteriormente na integração da descrição desses recursos a outros *datasets* universais e publicados no LOD.

O processo de conversão é explicado por Haslhofer e Isaac (2011), sendo sintetizado na Figura 1.

Na Figura 1 demonstra-se como os dados criados em EDM obtém as características apresentadas por tal modelo. O primeiro passo da conversão ocorre na obtenção dos metadados da base da Europeana. Na sequência, os metadados obtidos são convertidos em EDM, por meio de folhas de estilos, que realizam a conversão entre os dois modelos, transformando os metadados em EDM com URIs, representados no padrão RDF/XML. Estando com os metadados convertidos, inicia-se um dos processos mais importante para tornar efetivamente esta iniciativa em LOD, que é o enriquecimento dos dados.

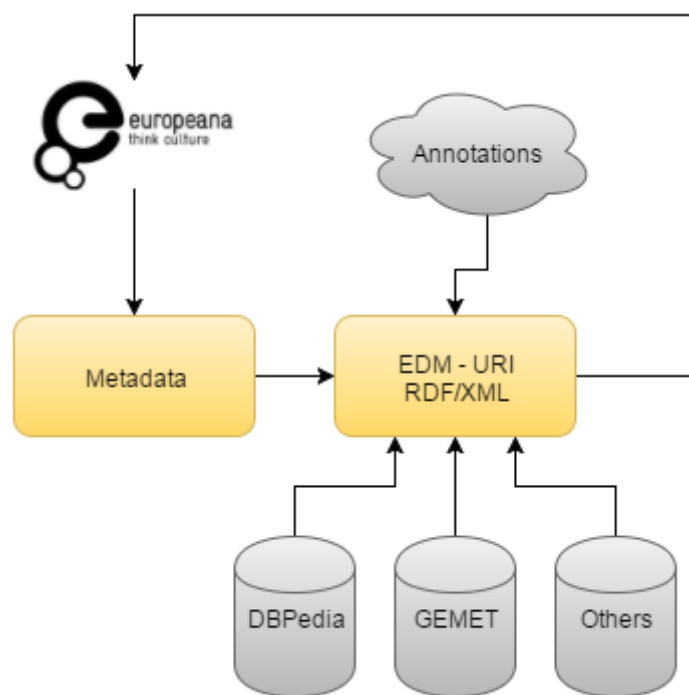


Figura 1: Conversão de Metadados para o EDM.

Fonte: Elaborado pelos autores.

Primeiramente, são inseridas anotações feitas pelo escritório da Europeana, que contém informações relevantes sobre os objetos culturais, e na sequência são relacionados os itens de tesouros das representações no GEMET (<https://www.eionet.europa.eu/gemet/>) e nos itens de pessoas, são feitos relacionamentos com o DBpedia (<http://dbpedia.org>). Esse processo é fundamental para que sejam feitas ligações com URIs já existentes, e que possua uma maior representatividade, tornando assim os dados ligados com outras bases.

Uma outra questão que insere confiabilidade nos dados contidos no EDM são os sete requisitos mais notáveis que guiaram a construção deste modelo. São eles: (i) distinção entre os objetos em si (livros, pinturas, esculturas) e as suas representações digitais; (ii) distinção entre o registro de metadados e o próprio objeto; (iii) é permitido a existência de registros distintos sobre um mesmo objeto, podendo conter declarações contraditórias sobre tal objeto; (iv) suportar objetos que são compostos por outros objetos; (v) compatibilidade com diferentes tipos de abstração de descrição; (vi) o EDM utiliza um formato padrão de metadados, que pode ser especializado; (vii) suportar recursos contextuais, o que inclui vocabulários controlados (EUROPEANA, 2013).

Os requisitos apresentados demonstram a preocupação do projeto em prover uma estrutura com semântica inserida. Este fato pode ser verificado ao indicar a distinção entre objetos em si das representações digitais, a inserção de diversos registros de um mesmo objeto, o suporte a vocabulários controlados. Neste sentido, uma outra característica dos dados do EDM que indicam existência de associações entre os objetos em si e as suas representações, é explicitada na existência de três tipos de classes principais de recursos:

“Objeto de património cultural” em si (uma pintura, um filme, uma partitura, um livro ...) (**edm:ProvidedCHO**); -uma ou mais representações digitais acessíveis deste objeto, algumas das quais serão utilizadas como pré-visualização (a imagem digital de uma pintura) (**edm:WebResource**); - uma agregação para representar o resultado da atividade realizada. (**ore:Aggregation**) (EUROPEANA, 2013, p.8, tradução nossa, grifo nosso).

As três classes principais do modelo EDM indica a importância que a realização de agregações e de associações possui neste esquema, ressaltando a necessidade de haver uma diferenciação entre os objetos em si e as representações gráficas, como descrito anteriormente em um dos requisitos do modelo EDM.

Santarem Segundo (2015) demonstra ainda que o modelo de dados EDM consegue expressar, em forma de relações, quando diversos objetos possuem algum tipo de ligação,

como “[...] ligações entre obras de um mesmo artista que estejam em locais diferentes, ou então expressões diferentes de uma mesma obra interligadas através das propriedades do EDM.” (SANTAREM SEGUNDO, 2015, p. 235). Com esta afirmação, o autor expressa um pouco das possibilidades que o EDM possui por estar conectado à diversas outras bases de dados do LOD e por interligar seus próprios dados.

A Figura 2 ilustra como uma representação em EDM se configura em esquema de grafo. Por meio dessa figura é possível identificar que o RDF do modelo explicita um grande número de relações, que favorecem a identificação do contexto e das características de um objeto cultural contido na Europeiaana.

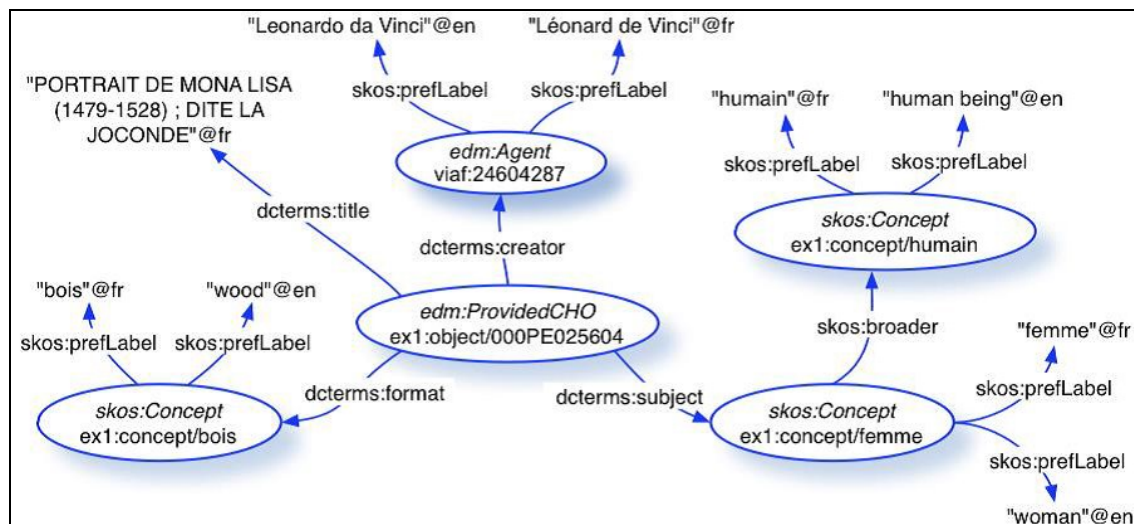


Figura 2: Representação Gráfica EDM.

Fonte: Europeiaana (2013, p. 15)

Na Figura 2, identifica-se a existência de informações de diversas fontes de dados se relacionando, como as informações da classe “edm:Agent” que utiliza os dados de autoridade do VIAF, número 24604287, que faz referência ao artista “Leonardo da Vinci”, além de utilizar informações de um vocabulário controlado construídos com a linguagem SKOS, bem como informações do próprio Museu do Louvre tratando do quadro da “Mona Lisa”. Uma outra característica é a existência de rótulos (*label*) indicando as informações da obra de arte, tanto na língua francesa quanto na inglesa, o que facilita uma melhor recuperação e apresentação posterior das informações.

Aprofundando o nível de relações existentes no EDM, identifica-se alguns relacionamentos semânticos existentes entre os objetos digitais. Neste sentido, os relacionamentos “edm:isRepresentationOf” e “edm:isDerivativeOf” possibilitam a descrição de um objeto digital que representa uma determinada obra e o relacionamento entre uma obra que é derivada de outra. Vale destacar a diferença existente entre uma representação e uma derivação no contexto do EDM, em que uma representação indica um objeto que represente com as mesmas características um objeto, como por exemplo uma foto digital de uma obra de arte, enquanto a derivação indica um objeto que foi construído com características distintas, mas inspirada em uma obra de arte, como um filme, por exemplo.

Esses dois tipos de relacionamentos promovem com que as representações dos dados estejam contextualizadas, pois deixará claro que uma determinada imagem digital ou uma derivação de uma obra fique vinculada a seu objeto original. Um exemplo bastante claro é ilustrado na Figura 3, em que um objeto de representação está ligado ao objeto principal (“Portrait de Mona Lisa”), além de haver um objeto derivado (“Mona Lisa – 2000”)

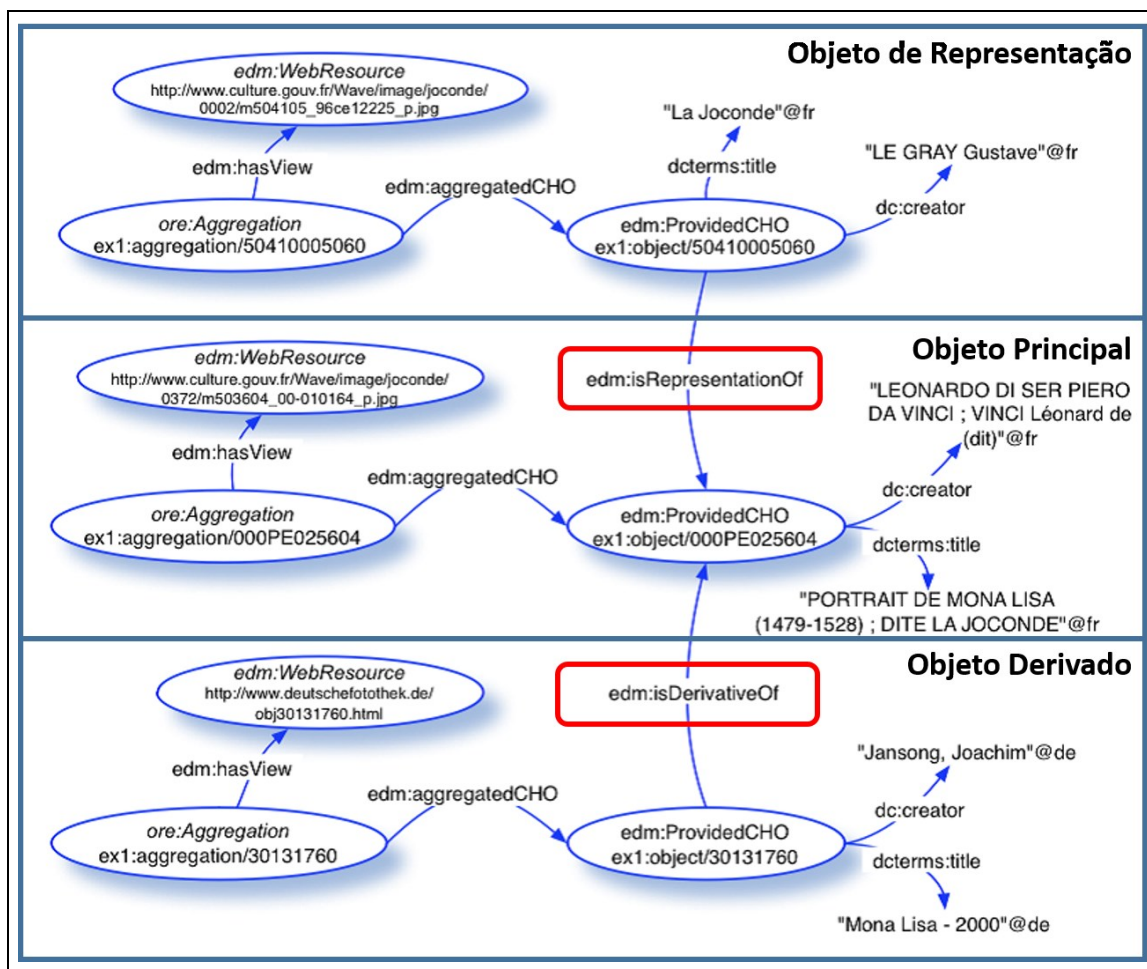


Figura 3: Representação de objetos vinculados.

Fonte: Adaptado Europeana (2013, p. 32)

Além desses relacionamentos, a documentação da Europeana indica a existência de outras relações como, “dcterms:hasPart” e “dcterms:isPartOf”, para representar a inserção de ligações entre objetos, e “edm:isRelatedTo”, que indica a existência de uma relação entre objetos, para que um usuário possa explorar outros objetos relacionados.

Uma consequência da utilização de grafos RDF para a disponibilização dos dados é a possibilidade de recuperação dos mesmos por meio do uso do SPARQL. Desta forma, o *dataset* da Europeana no LOD disponibiliza uma interface SPARQL *Endpoint*, que permite a realização de consultas nos dados estruturados. Na sequência é relatado um processo exploratório realizado na base de dados da Europeana, por meio do uso do SPARQL *Endpoint* deste serviço.

4 RESULTADOS: Um experimento realizado utilizando a linguagem SPARQL

A utilização do SPARQL na base de dados da Europeana permite com que sejam trabalhados com diversos tipos de relacionamentos e inferências. Tal fato ocorre principalmente, pois, por mais que exista interfaces visuais da Europeana para a realização de filtros, existe uma gama de possibilidades quando se utiliza as relações existentes em um modelo que tem como base o RDF.

Neste sentido, buscando compreender como ocorrem as relações dos objetos no padrão do EDM, realizamos algumas consultas no SPARQL *Endpoint* da Europeana¹. Por meio deste processo torna-se possível identificar os enriquecimentos que os dados sofreram, bem

¹ SPARQL End-Point Europeana. Disponível em: <<http://europeana.ontotext.com/sparql>>. Acesso em: 1º jun. 2016.

como compreender com maior propriedade como são as estruturas existentes no modelo EDM e na Europeia como um todo.

Primeiramente, buscamos verificar se era possível selecionar os registros vinculados a um determinado criador, utilizado para a recuperação vocabulários como o VIAF ou o DBpedia. Para isto, construímos uma consulta SPARQL que visava recuperar o título de um objeto, o provedor destes dados e o nome do criador, recuperando pelo registro de autoridade VIAF de número 24585191, que faz referência ao artista Michelangelo. Além disso, a consulta relaciona duas entidades relacionadas, por meio da propriedade “ore:proxyIn”, para identificar qual é o provedor das informações extraídas. Na Figura 4 apresenta-se um grafo, que demonstra na estrutura RDF as relações existentes entre os dados, que a consulta SPARQL explorou.

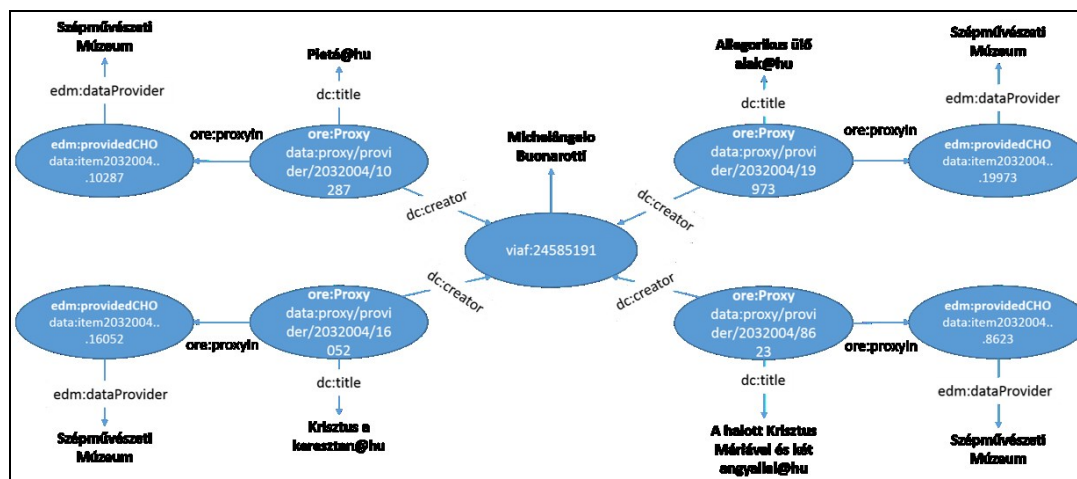


Figura 4: Consulta e resultado extraindo informações de Michelangelo.

Fonte: Elaborado pelos autores.

Neste grafo apresentado na Figura 4, é possível verificar que a entidade central é a criador, pois a mesma foi utilizada pelo SPARQL para recuperar as outras entidades. Outra característica que pode ser visualizada pelo grafo são os relacionamentos de entidades, em que há a relação entre o criador com a entidade “ore:proxy”, que por sua vez está relacionado com a entidade “edm:providedCHO”. A relação do tipo “ore:proxyIn” indica a existência de agregações de entidades dentro do EDM, em que as relações do tipo “ore:proxyIn” e “ore:proxyFor” tem a função de interligar duas entidades a uma terceira, contendo dados acerca de um objeto real que está representado por meio destas três entidades dentro do EDM. Em síntese, o uso da propriedade “ore:proxyIn” tem a função de possibilitar que dados de diversas fontes possam ser agregados e utilizados, como ocorre no exemplo indicado pela Figura 4.

A Figura 5 apresenta a consulta criada com os respectivos resultados obtidos.

<pre> PREFIX viaf: <http://viaf.org/viaf/> PREFIX dc: <http://purl.org/dc/elements/1.1/> PREFIX ore: <http://www.openarchives.org/ore/terms/> PREFIX edm: <http://www.europeana.eu/schemas/edm/> SELECT DISTINCT ?titulo ?provedor ?criador WHERE { ?a dc:creator viaf:24585191; dc:creator ?criador; ore:proxyIn ?agreg; dc:title ?titulo. ?agreg edm:dataProvider ?provedor.} ORDER BY DESC (?criador) LIMIT 4 </pre>	Título'	Provedor	Criador
	Pietá@hu	Szépművészeti Múzeum	Michelangelo Buonarroti Michelangelo Buonarroti Michelangelo
	Krisztus a kereszten@hu	Szépművészeti Múzeum	Michelangelo Buonarroti Michelangelo Buonarroti Michelangelo
	Allegorikus ülő alak (a négy temperatmentum egyike)@hu	Szépművészeti Múzeum	Michelangelo Buonarroti Michelangelo Buonarroti Michelangelo
	A halott Krisztus Máriával és két angyallal@hu	Szépművészeti Múzeum	Michelangelo Buonarroti Michelangelo Buonarroti Michelangelo

Figura 5: Consulta e resultado extraindo informações de Michelangelo.

Fonte: Elaborado pelos autores.

A parte da esquerda da Figura 5 apresenta a consulta SPARQL, havendo uma limitação de apresentar os quatro primeiros resultados e ordenar em ordem decrescente segundo o criador. Nos resultados obtidos, a direita da Figura 5, é possível identificar quatro

resultados, em que por exemplo, a primeira obra é a “Pietá”, descrita na língua húngara, tendo como o provedor desta informação o “Szépművészeti Múzeum” e como seu criador o artista “Michelangelo Buonarroti”, conforme indicado na consulta ao inserir a informação do VIAF que representava esse artista.

Desta forma, por meio do teste realizado é possível verificar que os dados disponibilizados no EDM utilizam os processos de agregação com bastante frequência e eficiência, ao passo que permite uma recuperação adequada pelo SPARQL.

Outras possibilidades que o SPARQL permite diz respeito a processos de agrupamento dos dados, que podem fornecer informações relevantes acerca dos dados, além da possibilidade de realizar associações com outras bases de dados do LOD. Desta forma, construímos duas consultas demonstradas nas Figura 6 e 7, com os seus respectivos resultados.

A primeira consulta, na Figura 6, apresenta um processo de agrupamento em que foram contabilizados a quantidade de objetos de acordo com o país dos mesmos, a resposta retornada desta consulta, apresenta como resultado que a Itália, Espanha, França e Bulgária são os países com maior número de elementos dentro da Europeia.

<pre> PREFIX viaf: <http://viaf.org/viaf/> PREFIX dc: <http://purl.org/dc/elements/1.1/> PREFIX data: <http://data.europeana.eu/> PREFIX ore: <http://www.openarchives.org/ore/terms/> PREFIX edm: <http://www.europeana.eu/schemas/edm/> SELECT ?c (COUNT(?agregado) as ?contagem) WHERE { ?agregado edm:country ?c.} GROUP BY ?c LIMIT 5 </pre>	país	contagem
	Italy	3530315 (xsd:integer)
	Spain	3910607 (xsd:integer)
	France	4172341 (xsd:integer)
	Bulgaria	95387 (xsd:integer)
	Croatia	65402 (xsd:integer)

Figura 6: Consulta SPARQL contagem dos países.

Fonte: Elaborado pelos autores.

A segunda consulta foi realizada o que é chamado de consultas federadas (*federated queries*), que apresenta a função de recuperar dados de diversos *datasets*, realizando relacionamentos entre eles, apresentada na Figura 7. Neste caso, a consulta visa recuperar os dados que tem como criador “dbpedia:Pablo_Picasso”, que é o recurso que diz respeito ao pintor Pablo Picasso na base de dados DBpedia, assim, buscamos relacionar os dados que estão das obras do Picasso, relacionando a informações de sua página no DBpedia.

<pre> PREFIX dc: <http://purl.org/dc/elements/1.1/> PREFIX dbpedia: <http://dbpedia.org/resource/> PREFIX dbpedia-owl: <http://dbpedia.org/ontology/> PREFIX dbpprop: <http://dbpedia.org/property/> SELECT ?obra ?criador_movimento ?criador_nome WHERE { ?a dc:creator dbpedia:Pablo_Picasso; dc:title ?obra. SERVICE <http://DBpedia.org/sparql> { SELECT ?criador_movimento ?criador_nome WHERE { dbpedia:Pablo_Picasso dbpedia-owl:movement ?criador_movimento; dbpprop:name ?criador_nome. } } }LIMIT 5 </pre>	obra	criador_movimento	criador_nome
	Denkmal für Guillaume Apollinaire@de	http://dbpedia.org/resource/Surrealism	Pablo Picasso@en
	recto: Anya gyermekével@hu	http://dbpedia.org/resource/Cubism	Pablo Picasso@en
	Szegényes teríték (Mutatványosok-sorozat)@hu	http://dbpedia.org/resource/Cubism	Pablo Picasso@en
	Lóitatás (Mutatványosok-sorozat)@hu	http://dbpedia.org/resource/Cubism	Pablo Picasso@en
	Mutatványosok (Mutatványosok-sorozat)@hu	http://dbpedia.org/resource/Cubism	Pablo Picasso@en

Figura 7: Consultas Federada no SPARQL contendo dados do DBpedia.

Fonte: Elaborado pelos autores.

Estas duas consultas demonstram o potencial que o EDM possui, quando utilizado com o SPARQL, em que é possível realizar agrupamento dos dados além de realizar consultas que exploram o EDM juntamente com outras bases de dados, como o DBpedia. Além disso, as três consultas demonstradas neste trabalho conseguem explicitar parte do potencial que o EDM possui, por meio dos relacionamentos entre os dados, a utilização de outras bases de dados e a utilização de vocabulários controlados.

4.1 Discussão

A Web apresenta um cenário em que a maioria das informações se encontram desestruturadas e com difícil acesso. Em contrapartida, o surgimento do LOD trouxe uma nova forma de organizar o conhecimento dentro da Web.

O esquema funcional do LOD contribui na publicação e utilização de dados devidamente estruturados que possibilitam melhores condições de recuperação. Hjørland (2008) ao discorrer sobre a organização do conhecimento, insere algumas características sobre a organização do conhecimento, como a organização social das mídias e a produção e difusão do conhecimento.

Por meio dessa afirmação, verifica-se que a área da Organização do Conhecimento possui grande influência em pesquisas e técnicas que tratem das estruturas dos dados, além de estudos que perpassam por áreas como a indexação, a representação e a recuperação da informação. Diante do exposto, verificamos que há uma profunda relação entre as novas formas com que a informação está sendo tratada dentro da Web, e os estudos da disciplina de Organização do Conhecimento. Assim, o cenário do desenvolvimento da Web, da Internet e do LOD, apresenta uma possibilidade de novos caminhos a serem trilhados dentro desta disciplina.

Dessa forma, o modelo de dados criado pela Europeana, EDM, acrescenta elementos que torna o seu acervo integralmente conectado ao ambiente Web, apresentando as características da Web Semântica e do LOD, ao mesmo tempo, que apoia outras iniciativas abertas, para utilizarem seus dados e suas definições na construção de novos conhecimentos.

O modelo de dados EDM, ao integrar diversos padrões, possibilita que os dados representados estejam disponibilizados de maneira acessível e organizada, tornando a Europeana uma referência no contexto do LOD e da Organização do Conhecimento.

A Europeana sendo uma biblioteca virtual que reúne conhecimentos de diversos séculos, pode ser um ponto de confiança e de partida para os diversos serviços dentro da Web, além de ser um exemplo de como a Organização do Conhecimento torna-se fundamental no contexto do LOD, sendo uma referência para os diversos outros serviços e instituições que desejam inserir seus dados na Web.

Nesse contexto, o EDM se mostra como um modelo que efetivamente possibilita uma representação adequada dos objetos culturais disponibilizadas por meio de uma biblioteca, ao se tratar do uso eficaz e eficiente das principais tecnologias da Web. Por mais que existam diversos outros modelos que se baseiam no LOD e nos conceitos da Web Semântica na descrição de metadados, como MODS RDF e Bibframe, o EDM surge como um modelo aplicado a uma das principais bibliotecas virtuais do globo, permitindo que os relacionamentos entre uma obra e suas representações e manifestações estejam descritos corretamente. O EDM liga os conteúdos que se relacionam, apresentando-se como um meio de organizar os conhecimentos vinculados a objetos digitais na área cultural, que se mostra sólido quando se verifica a abrangência e a importância que a Europeana vem ganhando ao passar dos anos.

Os testes utilizando o SPARQL, permitiu comprovar tais questões, ao demonstrar que o modelo EDM torna a representação dos objetos digitais recuperáveis pelas tecnologias de Web Semântica e LOD, contendo inclusive possibilidades que permitam a contextualização dos dados. Neste sentido, destaca-se o uso das bases de conhecimento GEMET e DBpedia, que contém informações estruturadas que são referenciadas por grande parte dos *datasets* do LOD. Essa integração com outras bases é essencial para tornar os dados universais e dentro de um contexto de interligação de fontes de dados.

Ainda por meio dos testes com o SPARQL, identificamos que o modelo de dados da Europeana utiliza com frequência propriedades de agregação, que une informações de diversos locais para tornar as informações mais enriquecidas. Vale destacar que o uso de tecnologias como o RDF para descrever os registros possibilitou a criação de consultas com o SPARQL, mostrando-se bastante efetivo para realizar agrupamentos, que permite contagens acerca das obras contidas nas bases de dados, além da possibilidade de realizar inferências.

Os dados do RDF apesar de apresentar alguns problemas, como diferentes padrões para realizar as descrições, como em alguns dados que se utiliza o VIAF para fazer referência ao criador e em outros momentos se utiliza o DBpedia, existe uma estrutura funcional, que permite a exploração adequada por parte dos usuários tanto humanos quanto não-humanos. Essa característica se mostra em consonância com a proposta do Linked Open Data de

Berners-Lee (2006), que como relatado, tem como um de seus objetivos tornar possível a descoberta de dados pelos usuários.

Assim, a representação da Europeana, por meio de seu modelo EDM, no LOD mostra-se como um modelo de referência, na representação das Humanidades Digitais efetivamente na Web, sendo um exemplo de como as bibliotecas, museus e arquivos, e a Ciência da Informação e a Organização do Conhecimento podem guiar suas pesquisas no intuito de identificar os melhores modelos de inserir corretamente os objetos digitais culturais no ciberespaço como um todo.

5 CONSIDERAÇÕES FINAIS

A proposta da Europeana mostra-se inovadora pelo caráter tanto da disponibilização quanto da preservação dos objetos culturais da Europa. Quando esse conjunto de dados se apresenta também como um *dataset* do LOD, transcende a barreira do ser humano e permite com que os objetos digitais estejam disponíveis também para acesso por agentes computacionais inteligentes.

Ao organizar a descrição de seus recursos ligando seus objetos a *datasets* públicos do LOD, o *dataset* da Europeana apresenta-se como um dos principais modelos de informação cultural de acesso público e semanticamente organizado disponíveis na Web.

Vale destacar que, as características apresentadas pelo modelo de dados EDM se mostram aderentes aos conceitos da disciplina de Organização do Conhecimento. Isto porque, na concepção do projeto da Europeana e do EDM, estruturou-se as informações baseadas nos conceitos de bibliotecas, arquivos e museus, seguindo uma organização que facilitasse uma posterior recuperação.

O processo de recuperação dos dados apresenta ainda, a facilidade de estar de acordo com o LOD, que possibilita a extração com técnicas de SPARQL. Esta característica aumenta a visibilidade do EDM, pois permite que outros serviços possam utilizar os dados da Europeana, tornando esta iniciativa uma referência para outros projetos de LOD.

As consultas SPARQL realizadas no presente trabalho, demonstraram o potencial que o uso dos dados aderentes as diretrizes do LOD contém, em que são possibilitados relacionamentos e associações com outras bases de dados. Além disso, o uso do SPARQL indica a aderência do modelo EDM na utilização dos conceitos e tecnologias da Web Semântica, permitindo a contextualização dos dados.

Neste contexto, destacamos ainda a relação existente entre as Humanidades Digitais e a Europeana, que apresenta uma convergência importante. A Europeana é uma das principais iniciativas que utilizam o conceito de Humanidades Digitais, pois permite uma integração entre os objetos culturais tradicionais, com suas representações em ambientes digitais e virtuais. A Europeana se mostra como a principal expoente neste âmbito, especialmente pela abrangência mundial que a mesma possui.

Dentro do contexto da dimensão aplicada da organização do conhecimento, a materialização e disponibilização dos objetos digitais da Europeana, descritos com o padrão EDM e ligados ao LOD, representa um ciclo completo de uso de conceitos e aplicações práticas que tem como resultado um produto para uso da sociedade civil, uma das funções da Ciência da Informação como Ciência Social Aplicada.

AGRADECIMENTOS

Agradecemos a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) número de processo 2015/01517-2 e a CAPES pelo auxílio ao trabalho.

REFERÊNCIAS

- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific American**, v. 284, n. 5, p. 28-37, 2001.
- BERNERS-LEE, T. et al. **Uniform Resource Identifier (URI): Generic Syntax**. 2005. Disponível em: <<https://tools.ietf.org/html/rfc3986#section-1.1>>. Acesso em: 16 mai. 2016.
- BERNERS-LEE, T. **Linked data principles**. 2006. Disponível em: <<https://www.w3.org/DesignIssues/LinkedData.html>> Acesso em: 10 abr. 2015.
- CYGANIAK, R.; JENTZSCH, A. **The Linking Open Data cloud diagram**. 2014. Disponível em: <<http://lod-cloud.net/>>. Acesso em: 13 ago. 2015.
- EUROPEANA. **Definition of the Europeana Data Model v5.2.6**. 2014. Disponível em: <http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM%20Definition%20v5.2.6_01032015.pdf> Acesso em: 17 fev. 2016.
- EUROPEANA. **Europeana Data Model Primer**. 2013. Disponível em: <http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf> Acesso em: 2 jun. 2016.
- FERREIRA, J. A.; SANTOS, P. L. V. A. C. O modelo de dados resource description framework (RDF) e o seu papel na descrição de recursos. **Informação & Sociedade: estudos**, João Pessoa, v. 23, n. 2, p. 13-23, maio/ago. 2013. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/15436/9681>> Acesso em: 16 maio 2016.
- HASLHOFER, B.; ISAAC, A. data. europeana. eu: The europeana linked open data pilot. In: **International Conference on Dublin Core and Metadata Applications**. 2011. p. 94-104.
- HEATH, T. **Linked data-connect distributed data across the web**. 2012. Disponível em: <<http://linkeddata.org/>>. Acesso em: 14 aug. 2015.
- HJØRLAND, B.. What is knowledge organization (KO)?. Knowledge organization. **International journal devoted to concept theory, classification, indexing and knowledge representation**, 2008.
- SANTAREM SEGUNDO, J. E. Web Semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas internacionais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v.8, n. 2, 2015.
- 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: GUIMARÃES, J. A. C.; DODEBEI, V. (Org.). **Organização do Conhecimento e Diversidade Cultural**. 1ed. Marília: Fundepe, v. 3, p. 224-233, 2015.
- WINER, D.; ROCHA, I. E. Europeana: um projeto de digitalização e democratização do patrimônio cultural europeu. **Patrimônio e Memória**, 2013.
- W3C. **W3C Opens Data on the Web with SPARQL**. 2008. Disponível em: <<http://www.w3.org/2007/12/sparql-pressrelease.html.en>>. Acesso em: 06 jun. 2016.

Editores do artigo: Adilson Luiz Pinto, Rafaela Paula Schmitz e Enrique Muriel-Torrado