

Web Semântica na Europeana: uma abordagem do padrão EDM

Luciana Candida Silva

Universidade Federal de Goiás, Brasil

José Eduardo Santarem Segundo

Universidade de São Paulo Brasil

Nuno Freire

Instituto de Engenharia de Sistemas e Computadores Investigação e Desenvolvimento, Lisboa, Portugal

REVIEW

Resumo

Objetivo. A Web Semântica e o Linked Data enfatizam a reutilização e a ligação de recursos ricamente descritos na Web. Estes princípios ajustam-se ao propósito do Modelo de Dados da Europeana (EDM) de utilizar informações de recursos existentes e apoiar o seu enriquecimento por meio do estabelecimento de novas relações entre eles. Assim, o objetivo deste estudo é descrever as relações semânticas inseridas nos elementos do EDM, destacando as vantagens de uso desse modelo para recuperação da informação na Web e, dessa forma, incentivar a adoção de metodologias semânticas em projetos brasileiros.

Método. Trata-se de uma pesquisa de natureza qualitativa do tipo descritiva-documental, baseada na família de documentos do EDM. Primeiramente o estudo identificou os conceitos e tecnologias da Web Semântica e Linked Data, em seguida, analisou o memorial descritivo do modelo de dados da Europeana. O EDM foi detalhado em seus princípios e desenvolvimento, com ênfase nos elementos semânticos que modelam e apoiam a funcionalidade da Europeana.

Resultados. Como resultado foram identificadas as possibilidades de conexão dos dados de diferentes instituições, de forma a enriquecer as informações dos registros de um determinado objeto do patrimônio cultural.

Conclusões. Este estudo mostrou, por meio das relações semânticas, que a estrutura semântica do EDM constitui-se em uma referência a ser seguida para publicação de dados de projetos nacionais no Linked Open Data, para garantir a interconexão cada vez maior de dados, aumentar a velocidade da circulação da informação entre os interessados e acelerar novas descobertas.

Palavras-chave:

Biblioteca digital Europeana; Ciência da Informação; Linked Data; Modelos de Dados da Europeana – EDM; Web Semântica.

Semantic Web in Europeana: an EDM standard approach

Abstract

Objective. The Semantic Web and Linked Data emphasize on the reuse and linking of richly described resources on the Web. This principles fits the purpose of the Europeana Data Model (EDM), which is to make use of existing resource information as well as to support their enrichment by establishing new relationships between them. The objective of this study is to describe the semantic relationships included in the elements of the EDM, as well as to highlight the advantages of using this model to represent and retrieve information digitized culture on the Web and thus encourage the adoption of semantic methodologies in Brazilian projects.

Method. This is a qualitative research, based on a descriptive-document method, regarding the family of EDM documents. First we have identified the concepts and technologies of the Semantic Web and Linked Data, then we analyze the descriptive memoir of the data model of Europeana. We detail the principles and development of EDM, with an emphasis on the semantic elements that support specific functions in Europeana.

Results. We have identified the possibilities of connecting data from different institutions in order to complement and enrich the information of the records about a particular cultural heritage object.

Conclusions. This study showed that the semantic structure of the EDM constitutes a reference to be followed for publication of data of national Linked Open Data projects to ensure the increasing interconnection of data, to increase the speed of the circulation of the information between the interested parties and accelerate new discoveries.

Keywords:

Europeana Data Model – EDM; Europeana digital library; Information Science; Linked Data; Semantic Web.

1 Introdução

A popularização da internet e o avanço das tecnologias da informação e da comunicação, a partir da década de 90, motivaram a disponibilização e o acesso da informação em meio digital. Entretanto, as informações são disponibilizadas na Web de modo a serem interpretadas, exclusivamente, por pessoas, prejudicando o resultado da busca da informação inserida em diferentes contextos. Assim, para obter o acesso rápido a informação relevante, é necessária a adoção de conceitos e tecnologias que permitam aos computadores o processo automático da organização semântica das informações. Neste contexto, destacam-se as iniciativas que seguem os princípios da Web Semântica para publicar e conectar dados no âmbito da Web.

A Web Semântica é a extensão da Web atual, em que são aplicadas tecnologias para atribuir à informação um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação (BERNERS-LEE; HENDLER; LASSILA, 2001). A Web Semântica vem evoluindo a fim de aprimorar o serviço de busca e melhorar a qualidade da recuperação dos dados na Web. Para tanto, o World Wide Web Consortium -W3C estabelece padrões para criação e interpretação dos dados na Web e dissemina tecnologias que vão se agregando em busca da Web Semântica ideal.

Em 2006, Tim Berners-Lee criou o conceito de Linked Data para se referir a um conjunto de regras a serem seguidas para possibilitar a interconexão, o reuso e o enriquecimento de dados na Web para viabilizar a implantação da Web Semântica.

A Web Semântica não é um projeto finalizado, mas é possível observar iniciativas que estão publicando seus conjuntos de dados com o uso de conceitos e tecnologias da Web Semântica e Linked Data. Entre as iniciativas, destaca-se a proposta do modelo de dados da Europeana, o Europeana Data Model - EDM, o qual está sendo construído com o uso de vocabulários reconhecidos internacionalmente, que possibilitam ao modelo uma riqueza capaz de acomodar normas de diferentes domínios. Destaca-se que o EDM é considerado uma referência na organização de objetos digitais de cultura, com vista à garantia do amplo acesso e utilização de materiais de domínio público digitalizado. Além disso, o EDM é o padrão de descrição de recursos utilizado para disponibilizar o conjunto de dados da Europeana no Linked Open Data - LOD de forma a serem livremente utilizados, reutilizados e redistribuídos por qualquer pessoa.

Diante da relevância de iniciativas que contemplam a publicação de dados abertos e semânticos para busca e recuperação da informação de qualidade na Web, esta pesquisa visa responder às seguintes questões de pesquisa: Como se dão as relações semânticas entre os elementos do EDM? Quais as vantagens de uso do padrão EDM para a recuperação da informação na Web? A explicitação dessas relações e das suas vantagens de uso poderá incentivar a adoção de metodologias semânticas em projetos brasileiros que visem a interconexão entre os dados de diferentes domínios e o enriquecimento de seus datasets, permitindo aos usuários obterem resultados mais precisos.

2 Web Semântica e Linked Data

A Web Semântica visa atribuir à informação um significado bem definido por meio da aplicação de conceitos e tecnologias utilizados pelos computadores para desempenhar operações que darão significado às palavras no âmbito da internet. A partir desse entendimento, Berners-Lee, Hendler e Lassila (2001) destacam que na Web Semântica os computadores não apenas serão capazes de apresentar a informação contida nas páginas Web, mas também compreendê-las.

Para Berners-Lee, Hendler e Lassila (2001), a estrutura de significados para os conteúdos das páginas da Web define um ambiente onde os agentes computacionais possam processar, compartilhar, reusar e enriquecer dados na Web, e assim realizar tarefas sofisticadas aos usuários.

Dessa forma, Berners-Lee (2006, tradução nossa), afirma que “a Web Semântica não consiste apenas em colocar dados na Web. Trata-se de fazer links, para que uma pessoa ou máquina possa explorar as redes de dados”.

Desde a proposta inicial, em 2001, a Web Semântica vem evoluindo para tornar possível a implantação de ambientes semânticos. Em sua fase inicial, a Web Semântica foi construída a partir de um modelo em camadas, com algumas tecnologias que constituíram a sua base estrutural. Após a publicação do artigo *The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, em 2001, de autoria de Tim Berners-Lee, James Hendler e Ora Lassila, vários estudos e publicações em torno do tema proporcionaram uma reavaliação das camadas, surgindo novos formatos para representar os dados na Web. Como

parte desse desenvolvimento, surgiu o conceito de Linked Data como uma maneira de publicar dados conectados na Web.

O Linked Data segue os princípios definidos em 2006 por Tim Berners-Lee para criar ligações entre recursos de distintas fontes e promover a reutilização e o enriquecimento de dados na Web. Estes princípios são:

1. Usar Uniform Resource Identifier - URIs para nomear itens;
2. Usar URIs HTTP para que pessoas possam procurar esses nomes;
3. Ao consultar uma URI, fornecer informações em formatos padronizados úteis [Resource Description Framework - RDF, SPARQL etc.];
4. Incluir sentenças com links para outras URIs, para permitir que itens relacionados possam ser descobertos (BERNERS-LEE, 2006).

Assim, Bizer, Heath e Berners-Lee (2009, p. 2, tradução nossa), definem o Linked Data como “um conjunto de boas práticas para publicar e conectar dados estruturados na Web, com o intuito de criar uma Web de dados”. As boas práticas adotam padrões reconhecidos internacionalmente e recomendados pelo W3C para construir a Web de Dados.

A principal tecnologia recomendada pelo W3C para o desenvolvimento de projetos semânticos é a RDF usada para descrever os recursos e acessar dados de diversas fontes no âmbito global. Para Berners-Lee, Lassila e Hendler (2001), a linguagem RDF expressa o significado das etiquetas, o qual tem como princípio fornecer interoperabilidade aos dados, de forma que possa contribuir com a recuperação de informações de recursos na Web. Para descrever a relação entre recursos e codificar o significado das etiquetas, o modelo RDF oferece uma estrutura de triplas do tipo sujeito, predicado e objeto. O sujeito e o objeto representam dois recursos que são relacionados por um predicado. Por exemplo, Leonardo da Vinci (sujeito) é criador (predicado) da obra Mona Lisa (objeto). Os recursos são quaisquer coisas disponíveis na Web identificadas por URIs. Assim, as três partes da tripla são vinculadas através do uso de URIs. Segundo Isotani e Bittencourt (2015), os URIs proveem uma maneira única para identificar recursos e expressar relações entre eles.

O modelo de dados gráficos RDF pode ser apresentado em vários formatos. Entre eles aparecem o RDF in Attributes - RDFa, RDF eXtensible Markup Language - RDF/XML, N-Triples e Turtle. Além das serializações que suportam múltiplos grafos, como o JSON-LD, TRING e N-Quads. Para Isotani e Bittercourt (2015), as serializações permitem adicionar outros grafos, que se conectam ao grafo original, proporcionando a formação de datasets conectados.

Para estabelecer relações entre termos, a Web Semântica faz uso de ontologias. Segundo Patrício (2012), as ontologias são compostas por uma taxonomia, que define as classes de objetos e as relações entre eles, e por regras de inferência as quais permitem que os programas computacionais manipulem os termos de forma mais eficiente do que os seres humanos. As ontologias podem ser expressas em RDF Schema - RDFS ou pela Web Ontology Language - OWL. O RDFS fornece um sistema básico de classes e propriedades e indica como são usados em conjunto. A OWL é uma ontologia mais extensa e expressiva do que a RDFS, usada para descrever e definir termos dentro de um determinado domínio de interesse ou assunto e para descrever e definir as relações entre eles. Outras ontologias desenvolvidas e consolidadas são o Friend of a Friend - FOAF e o Simple Knowledge Organization System - SKOS. O FOAF tem por objetivo descrever relacionamentos entre pessoas e informações na Web. O SKOS desenvolve especificações e padrões para utilização de organização do conhecimento, como tesouros, esquemas de classificação, taxonomias e cabeçalhos de assuntos no âmbito da Web Semântica, com capacidade de expressar suas relações hierárquicas e associativas (RYAN; GRANT; COLLINS; STEFAN; LOPES, 2017). Para consultas e acesso a conjuntos de dados descritos em RDF, destaca-se a interface SPARQL. De acordo com a W3C (2013), o SPARQL refere-se a um conjunto de especificações que fornece linguagens e protocolos para recuperar e manipular dados armazenados em RDF.

A adoção dessas tecnologias, seguindo as melhores práticas do Linked Data, alavanca a infraestrutura da Web atual com ligações e compartilhamentos de dados estruturados para consumo humano e de máquinas, permitindo o relacionamento entre recursos de diferentes domínios. Esses relacionamentos criam uma rica rede de informações interconectadas na Web, favorecendo os resultados das buscas.

Segundo Isotani e Bittencourt (2015), os dados conectados, entretanto, não necessariamente precisam ser abertos. Por exemplo, uma entidade privada pode conectar dados, mas não deixá-los abertos. Os autores salientam que, apesar de existirem iniciativas em dados conectados de forma fechada, muitas iniciativas estão se preocupando com a conexão e publicações de dados de forma aberta. Para a Open Knowledge International (2004), o

Linked Open Data - LOD são dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa, estando sujeito, no máximo, à exigência de citar a fonte original e compartilhar com as mesmas licenças em que as informações foram inicialmente apresentadas.

Para Santarem Segundo (2015) e Coneglian & Santarem Segundo (2017) a iniciativa mais sólida no emprego dos recursos da Web Semântica com efetividade são os datasets publicados por meio do LOD, o qual conta com uma grande quantidade de bases de dados estruturadas e conectadas e vem crescendo ao longo dos anos. Nesse contexto, destaca-se a Europeana, um dos projetos de maior destaque no LOD, que por meio do padrão EDM publica dados em grande volume no contexto cultural e é destaque das Humanidades Digitais, área que integra a aplicação de tecnologias às ciências humanas.

3 Europeana

A Europeana é uma biblioteca digital que surgiu do projeto de tornar o patrimônio cultural da Europa disponível na internet, para garantir o acesso e uso do material digitalizado e para preservá-lo para as gerações futuras. A Europeana teve o seu ponto de partida no ano de 2000, quando o Projeto Gateway and Bridge to Europe's National Libraries - GABRIEL estabeleceu um portal com informações de coleções de 43 bibliotecas nacionais europeias. A Comissão Europeia co-financiou projetos de investigação e promoveu uma melhor colaboração entre Estados-Membros que tinham interesse em colocar em linha o seu patrimônio cultural. Este financiamento promoveu o crescimento do projeto GABRIEL para a The European Library - TEL, um mecanismo de busca e um centro de dados abertos para coleções de bibliotecas. O site da TEL foi lançado em 2005 e expandiu com a integração de mais bibliotecas nacionais e passou a ter maior capacidade de padronização e pesquisa. Esse crescimento ampliou a demanda por apoios e investimentos (JEFCOATE, 2006).

Em 2005 o presidente da França, Jacques Chirac, em conjunto com os seis primeiros ministros europeus, enviou uma carta ao presidente da Comissão Europeia, José Manuel Durão Barroso, apresentando as suas preocupações em preservar o patrimônio cultural europeu, sugerindo a criação de uma biblioteca virtual para acomodar o registro cultural científico da Europa, caso contrário a herança do patrimônio europeu, considerado inigualável em riqueza e diversidade, poderia perder o seu espaço na futura geografia do conhecimento. No mesmo ano, o presidente da Comissão Europeia saudou a proposta das bibliotecas digitais e adotou a estratégia i2010¹ para delinear a visão da iniciativa de bibliotecas digitais, a partir das ações de digitalização de coleções analógicas, acessibilidade online e conservação e armazenamento (JEFCOATE, 2006). Segundo Winer e Rocha (2013), a iniciativa da Biblioteca Digital Europeia foi lançada pela Comissão Europeia, ainda em 2005, e efetivada com o apoio das bibliotecas nacionais, outros organismos culturais dos Estados-Membros e do Parlamento Europeu.

O protótipo da Biblioteca Digital Europeia foi inaugurado oficialmente no dia 20 de novembro de 2008, no Palais de Charles de Lorraine, na Biblioteca Real Nacional de Bruxelas e disponibilizada ao público com aproximadamente 2 (dois) milhões de documentos digitalizados do acervo das bibliotecas e instituições culturais dos 27 Estados-Membros da União Europeia. A expectativa era a de disponibilizar em 2010 o acesso a 10 milhões de objetos representativos da riqueza da diversidade cultural da Europa. Atendida a expectativa, a partir de 2015 a Europeana reuniu em um único espaço digital o seu mecanismo de busca, exposições e blog e foi relançada como Europeana Collections. Atualmente, as Coleções Europeanas disponibilizam ao público um acervo com o diversificado patrimônio cultural e científico da Europa, composto por vários tipos de documentos digitalizados como livros, manuscritos, jornais, revistas, mapas, fotografias, pinturas, músicas, vídeos, documentários e programas de TV.

Considerando o avanço da Biblioteca Digital Europeia, hoje denominada de Coleções Europeanas, a TEL foi descontinuada, desde o dia 31 de dezembro de 2016, e o acesso aos seus dados passou a ser via Europeana. As Coleções Europeanas, comumente chamadas de Europeana, são dirigidas pela Fundação para a Biblioteca Digital Europeia, que reúne as principais associações de bibliotecas, arquivos, museus, arquivos audiovisuais e instituições culturais da Europa. A Europeana encontra-se instalada na biblioteca nacional neerlandesa, a Koninklijke Bibliotheek. Segundo Winer e Rocha (2013), a Europeana vem ampliando a democracia do conhecimento e motivando a criação de bibliotecas virtuais.

Além disso, a Europeana destaca-se como uma iniciativa na disponibilização de dados relacionados ao conceito das Humanidades Digitais. Como menciona Santarem Segundo (2015), é uma área em expansão que tem como uma de suas premissas a associação de técnicas computacionais aos produtos gerados pelas ciências humanas. O foco das humanidades digitais, neste contexto, é transferir os recursos tradicionais para os meios digitais, para potencializar o uso desses recursos.

A Europeana se desenvolve a partir do recebimento de um vasto e crescente número de objetos do patrimônio

cultural europeu, que são digitalizados por uma variedade de fornecedores de dados dos setores de arquivos, bibliotecas, museus e galerias de coleções audiovisuais de mais de 36 países diferentes. Estes fornecedores usam diferentes padrões de metadados para descreverem seus objetos, e a Europeana busca padronizar os dados para serem reutilizados com o menor custo possível e evitar a duplicação de esforços.

Para apresentar os dados de forma significativa em um contexto multicultural e multilíngue utilizando instrumentos semânticos do patrimônio cultural, a Europeana desenvolveu - em conjunto com especialistas técnicos das áreas de arquivos, bibliotecas, museus e das coleções audiovisuais -, o Europeana Data Model - EDM, o qual constitui em uma proposta de modelo para estruturar os dados que a Europeana irá gerenciar e publicar (EDM FACTSHEET, 2013?). A estrutura de funcionamento deste modelo permite a reutilização e a ligação de recursos descritos na Web e, portanto, permite uma experiência de usuário mais rica, conforme será descrito nos itens seguintes.

4 Modelo de Dados da Europeana – EDM

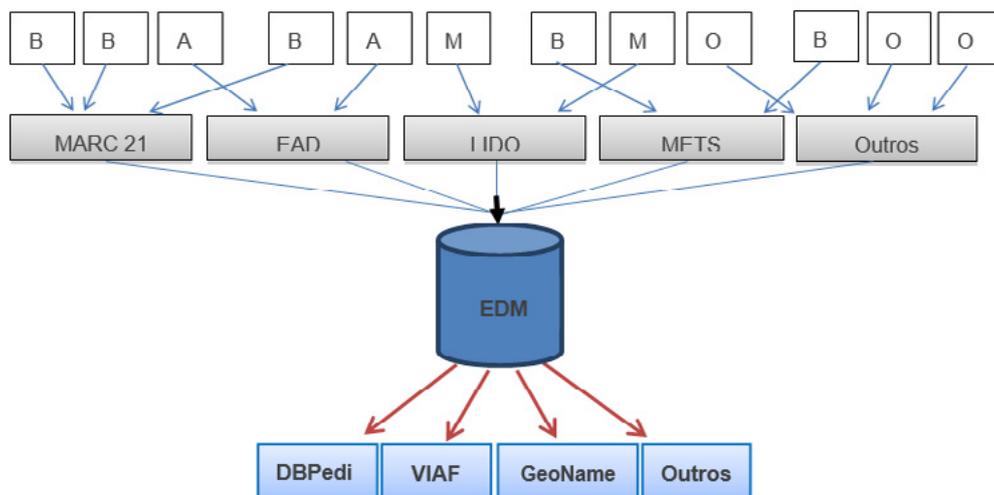
O modelo de dados da Europeana foi desenvolvido por meio de padrões e técnicas de domínio público, com base na Web Semântica, com capacidade para acomodar e enriquecer dados descritos em diferentes domínios, como nos padrões Lightweight Information Describing Objects - LIDO para museus, Encoded Archival Description - EAD para arquivos ou Metadata Encoding and Transmission Standard - METS para bibliotecas digitais (EDM PRIMER, 2013). A ideia é organizar os metadados de diferentes domínios e disponibilizar em um único ambiente digital para alcançar um público mais amplo.

O EDM é considerado uma melhoria do Europeana Semantic Elements - ESE, primeiro modelo de dados da Europeana. O modelo ESE trata-se de um perfil de aplicação baseado nos elementos Dublin Core - DC fornecendo um conjunto mínimo de dados que podem ser aplicados a materiais heterogêneos. O ESE produz um registro plano, isto é, agrega em um único registro de metadados, campos que se podem aplicar a diferentes entidades, onde nem sempre é possível identificar se um valor se aplica ao objeto original ou à sua representação digital, ou seja, há perda de riqueza dos dados originais. Atualmente ainda existem colaboradores que fornecem dados no padrão ESE, os quais são migrados para EDM, mas nem sempre apresentam os melhores resultados. O EDM é considerado um modelo mais flexível e permite criar um mapeamento baseado em um conjunto mais extenso de elementos do que era possível com o ESE. O mapeamento realizado diretamente para o EDM permite fornecer uma representação otimizada de seus dados de origem, além de apresentar um conjunto de metadados mais rico e com granularidade mais fina, garantindo um nível mínimo de interoperabilidade entre os dados de diferentes fornecedores (DOERR; GRADMANN; HENNICKE; ISAAC; MEGHINI; SOMPEL, 2010).

De acordo com Patrício (2012), o EDM visa suportar a representação de objetos mais complexos ou de estrutura hierárquica, como por exemplo, partes de um livro, arquivos, etc. Nessa abordagem, o EDM Primer (2013), exemplifica que em um livro digitalizado, os capítulos individuais, as ilustrações e o índice podem ser compreendidos individualmente e coletivamente. Os itens constitutivos como cartas e manuscritos podem ser igualmente compreendidos. Nesse sentido, pode-se dizer que o EDM permite melhorar qualitativamente os processos de como a Europeana recebe, gera e publica os dados agregados.

O formato de organização dos dados da Europeana, por meio do EDM, torna possível a publicação dos dados no LOD. Para manipular e exibir igualmente os metadados e disponibilizá-los como Linked Open Data, a Europeana realiza o mapeamento dos diferentes padrões de metadados recebidos dos provedores de dados dos setores de bibliotecas – MARC 21, arquivos - EAD, museus – LIDO dentre outros. Em seguida, os metadados são convertidos para o padrão EDM, por meio de folhas de estilos, que realizam a conversão entre dois modelos, transformando em EDM com URIs, representados no padrão RDF/XML. Estando com os metadados convertidos, inicia-se o processo de enriquecimento dos dados por meio de ligações com fontes externas como o DBPedia, VIAF, GeoNames e outros vocabulários de valores, como apresentado na figura 1.

Figura 1 – Conversão dos padrões de metadados em EDM



Fonte: Adaptado de Charles (2017)

Observa-se que o EDM é um padrão agregador na medida em que recebe dados de diferentes domínios, transforma-os no padrão da Europeana, faz os enriquecimentos com datasets externos, normaliza os valores dos dados e disponibiliza em um único ambiente digital com vista a favorecer o serviço de descoberta e a reutilização de recursos.

Para tanto, a construção do desenho do modelo EDM seguiu requisitos que o permitissem representar diferentes perspectivas sobre um objeto cultural, objetos complexos e informação de contexto. Os principais requisitos incluem:

- Distinção entre objetos fornecidos (pintura, livro, filme, site de arqueologia, arquivo arquivístico, etc.) e suas representações digitais.
- Distinção entre um objeto e o registro de metadados que o descreve.
- Permissão de ingestão de registros múltiplos para um mesmo item, o que pode conter declarações contraditórias sobre o mesmo objeto.
- Suporte para objetos que são compostos de outros objetos.
- Compatibilidade com diferentes níveis de abstração de descrição (por exemplo, se um fornecedor deseja apresentar descrições que seguem as distinções introduzidas no Grupo 1 do FRBR).
- Fornecimento de um formato de metadados extensível.
- Suporte para recursos contextuais, incluindo conceitos de vocabulários controlados. (EDM PRIMER, 2013, p.8 tradução nossa).

Além disso, uma motivação para o EDM é apoiar a integração dos vários modelos utilizados em dados do patrimônio cultural, de modo que todas as descrições originais possam ser coletadas e conectadas com alto nível de qualidade. Esta motivação define três princípios fundamentais de sua concepção: “permitir a integração de dados em um ambiente aberto; permitir a funcionalidade rica, possivelmente através de extensões; e reutilizar os padrões existentes sempre que possível” (EDM PRIMER, 2013, p. 8, tradução nossa).

Os princípios do EDM constituem a base para a escolha dos elementos semânticos para descrever formalmente as informações referentes aos objetos culturais, no âmbito da Web.

4.1 Elementos Semânticos do EDM

Os elementos do EDM são constituídos por vocabulários próprios, chamados de classes e propriedades, mas também, reutilizam vocabulários reconhecidos e indicados pelo W3C, para representar os dados e apoiar a funcionalidade da Europeana.

Os principais vocabulários reutilizados pelo EDM são o RDF e RDF(S), OWL para EDM, Open Archives Initiative Object Reuse and Exchange - OAI-ORE, SKOS, Dublin Core, W3C Data Catalog Vocabulary - DCAT e a Creative Commons Zero - CC0. Estes vocabulários permitem a reutilização e articulação flexível dos modelos existentes (EDM Primer 2013), “[...] o que torna o EDM um rico modelo de descrição dos objetos da Europeana, fortalecendo a estrutura de descrição semântica dos recursos” (CONEGLIAN; SANTAREM SEGUNDO, 2017, p. 4).

A utilização dos vocabulários permite que diferentes participantes estruturarem seus dados de uma forma que se adap-

tem aos seus dados originais e suas funções desejadas. Os provedores de dados podem criar conjuntos de dados simples ou mais complexos, dependendo da estrutura de seus dados de origem. Da mesma forma, a Europeana manipulará os dados internamente para executar suas funções de agregação e enriquecimento (EDM PRIMER, 2013). Segundo Patrício (2012), o EDM adota o vocabulário OAI-ORE para suportar a integração dos diferentes modelos usados para dados no âmbito do patrimônio cultural, a definição das entidades e das relações a que os metadados se referem. O modelo OAI-ORE define padrões para a descrição e troca de agregações de recursos da Web.

A estrutura de funcionamento do EDM é definida a partir do reuso e integração desses vocabulários às classes e propriedades para criação de um núcleo de descrição de cada objeto. As três classes principais que representam o objeto do patrimônio cultural são:

- One or more accessible digital of this object (edm:WebResource) - representa uma ou mais representações digitais do patrimônio cultural fornecido, algumas das quais serão utilizadas como pré-visualização (EDM PRIMER, 2013).
- One aggregation to represent the result of this provider's activity (ore:Aggregation) - cada agregação representa o conjunto dos elementos descritos do item, criado pelo fornecedor de dados ou pela Europeana. Uma agregação pode ser vista como a contribuição de um provedor para um objeto, o contexto digital que ele cria para esse objeto (EDM PRIMER, 2013).

De acordo com o EDM Primer (2013), as duas primeiras classes permitem captar a distinção entre obras que, se espera, sejam o foco do interesse dos usuários, e suas representações digitais, que são os elementos manipulados em sistemas de informação como a Europeana. Já a terceira classe, seguindo a abordagem ORE, demonstra que o objeto fornecido, juntamente com as representações digitais de um provedor de dados da Europeana, pode ser considerado como um todo lógico.

As diretrizes de mapeamento do EDM exemplificam o uso das classes para separar o objeto do patrimônio cultural de sua representação digital, para que os valores dos metadados sejam associados apropriadamente. Por exemplo, a pintura de Mona Lisa é representada pelo edm:ProvidedCHO e sua imagem digital pela classe edm:WebResource. Isso permite que as propriedades de metadados relevantes sejam aplicadas para cada classe. No exemplo da Mona Lisa, o edm:ProvidedCHO poderia ter uma propriedade dc:creator com o valor "Leonardo da Vinci" e o edm:WebResource a dc:format propriedade com o valor "jpg" (EDM MAPPING GUIDELINES, 2016). Assim como os metadados sobre um objeto e suas representações digitais são separados entre essas duas classes, há também um mecanismo para associar as classes relacionadas. Este mecanismo é o ore:Classe de agregação e é o objeto principal entre o edm:ProvidedCHO e o edm:WebResource(s). Ele tem propriedades para permitir a vinculação entre as classes associadas e também tem algumas propriedades descritivas mais familiares, como edm:dataProvider, que se aplicam a todo o grupo. Patrício (2012) expõe que as entidades definidas no modelo EDM são, pois, o próprio objeto do patrimônio cultural, as representações digitais desse objeto e as agregações das várias descrições do objeto.

Para enriquecer a descrição dos objetos representados pelas classes principais, o EDM possui em sua constituição as classes contextuais. As classes contextuais exploram a riqueza dos dados construídos por muitos provedores que usam arquivos de autoridade, vocabulários controlados e tesouros. A descrição dos elementos de um objeto do patrimônio cultural por meio das classes contextuais permite aos usuários navegarem na Europeana e obterem respostas às questões como Quem?, Onde?, Quando?, O que? e a licença de uso (EDM FACTSHEET, 2013?). As principais classes contextuais incluem: edm:Agent – who (quem), edm:Place – where (onde), edm:TimeSpan – when (quando), skos:Concept – what (o que) e cc:License – access and usage (licença de uso). Para cada classe principal e contextual é apresentado um conjunto de propriedades para a descrição e interconexão semântica entre recursos na Europeana (EDM MAPPING GUIDELINES, 2016).

O modelo EDM permite, por intermédio das agregações, capturar uma descrição do ambiente digital de um objeto submetido à Europeana e anexar informações descritivas aos vários recursos que participam deste ambiente, como é apresentado no contexto dos metadados descritivos.

Os metadados descritivos referem-se ao pacote de dados sobre um objeto do patrimônio cultural, ou seja, metadados para ProvidedCHO, WebResource, e agregação de quaisquer classes associadas (EDM MAPPING GUIDELINES, 2016). Estes metadados podem ser representados para o objeto fornecido, por exemplo, o criador. O criador (dc:creator ou dcterms:creator) é uma propriedade da classe principal edm:ProvidedCHO, que pode ser utilizada com o esquema Dublin Core para representar a pessoa responsável pelo objeto. Sendo assim, para representar as descrições do objeto, o EDM usa propriedades dedicadas que ele introduz ou reutiliza, como o edm:hasMet, dcterms:creator ou qualquer outra propriedade que os provedores julguem relevantes para descrever as características do objeto (EDM PRIMER, 2013).

4.2 Abordagens Descritivas do EDM

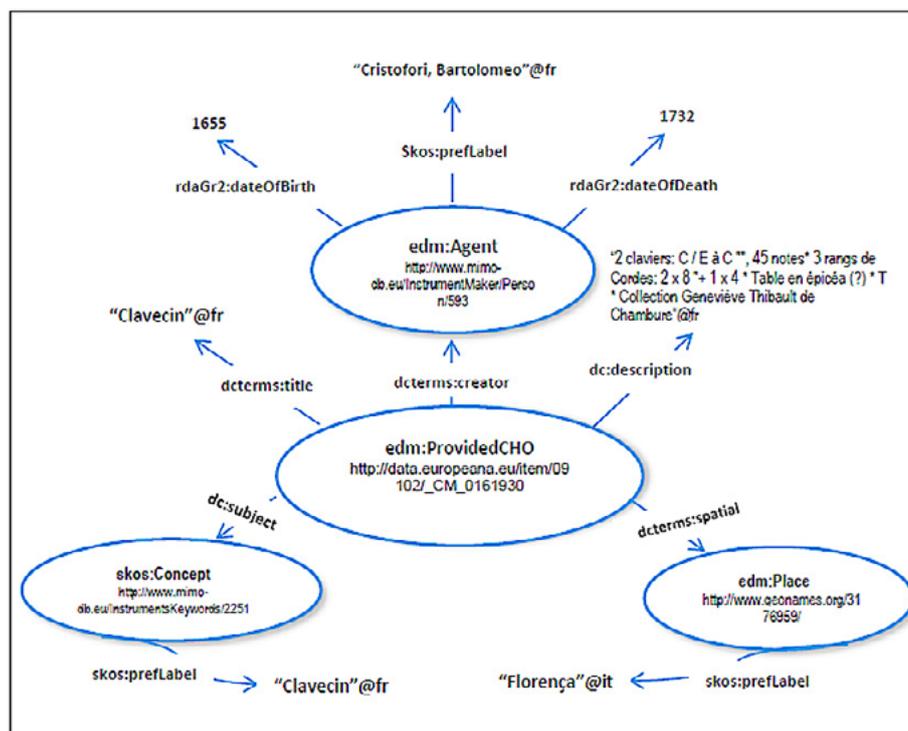
Os metadados descritivos agregados pelo EDM podem seguir as abordagens nas descrições centradas no objeto, nas descrições contextuais e nas descrições centradas em eventos, conforme apresentação do EDM Primer (2013):

- As descrições centradas no objeto são as propriedades descritivas do objeto, sejam elas strings simples ou recursos mais complexos denotando entidades do mundo real. Para descrever tais propriedades podem ser utilizadas propriedades Dublin Core (dc ou dcterms) como subpropriedades de propriedades EDM.
- As descrições contextuais descrevem valores nos metadados descritivos não contidos no objeto em si, mas em outro recurso na descrição. Isto significa que o EDM permite capturar valores como, por exemplo, a data de nascimento e de morte do autor a partir de outro recurso e enriquecer semanticamente com entidades contextuais. Para apoiar a modelagem desse enriquecimento semântico, o EDM adota as classes dedicadas à representação de entidades contextuais: edm:Agent para representação de pessoas e organizações; edm:Event para eventos; edm:Place para entidades espaciais; edm:TimeSpan, para períodos de tempo ou datas; skos:Concept, para todas as entidades de sistemas de organização do conhecimento como tesouros, esquemas de classificação e outras listas de autoridades. Essas classes permitem ligar URIs de fontes externas para adicionar informações aos dados existentes.
- As descrições centradas em eventos consideram que as descrições de objetos devem caracterizar os vários eventos nos quais os objetos foram envolvidos. Utiliza-se a classe edm: Event e as relações são representadas pelas propriedades: edm: wasPresentAt para ligar o recurso a um evento em que o mesmo esteve envolvido; edm: happenAt, para realizar a ligação entre o recurso e um local; edm: occurAt, para fazer a ligação entre o recurso e um período de tempo.

Segundo o EDM Primer (2013), com a abordagem RDF, o EDM permite que qualquer rede seja anexada a um objeto fornecido. A partir disso, entende-se que o EDM permite que abordagens centradas em objetos e centradas em eventos possam coexistir para o mesmo objeto. Assim, uma descrição centrada em eventos complementa a descrição do objeto, enriquecendo as entidades com os eventos que constituem todo o contexto de um objeto.

Para exemplificar a aplicação das classes e propriedades que compõem a estrutura do modelo EDM, apresenta-se a seguir uma representação gráfica baseada em informações do objeto Clavecin, de Bartolomeu Cristofori, exemplo extraído dos documentos das Coleções Europeanas.

Figura 2 – Representação gráfica da Estrutura EDM



Fonte: Elaborada a partir das Coleções Europeanas (2018)

A análise dos dados apresentados na representação gráfica da figura 2 permite destacar os seguintes pontos:

- Trata-se de uma abordagem centrada no objeto, na medida em que o objeto cultural fornecido (CHO) está relacionado diretamente a todas as suas características e enriquecido com as entidades contextuais.
- O objeto fornecido está representado pela classe `edm:ProvidedCHO`, que corresponde ao instrumento musical Clavecin, termo francês, registrado nas Coleções Europeanas sob o código identificador `CM_0161930`. O código identificador está inserido no próprio URI do objeto. A partir desse item foram realizadas ligações para outros dados desse mesmo objeto, ampliando a descrição e favorecendo a recuperação.
- O metadado descritivo `dcterms:creator` ligou o objeto ao artista responsável pelo instrumento musical, representado pela string “Cristofori, Bartolomeo”@fr. Para realizar essa ligação, o EDM adotou a classe `edm:Agent` (representação da pessoa), que utilizou dados de autoridade do Musical Instrument Museums Online (MIMO) para atribuir o valor ao recurso, fazendo referência ao artista “Bartolomeo Cristofori”. Para indicar a forma preferida do nome do agente na língua francesa, adotou a propriedade `skos:prefLabel`, enquanto que para informar as datas de nascimento e morte do agente, utilizou as propriedades `rdaGr2:dateOfBirth` e `rdaGr2:`, proporcionando mais informações para busca.
- Por meio da vinculação do objeto ao MIMO, via URI, é possível obter informações sobre as diferentes formas de apresentação do nome do criador, bem como as datas de nascimento e morte desse agente. Essas informações são apresentadas pelo MIMO em formato legível por usuários humanos e máquinas.
- Os metadados descritivos `dcterms:title` e `dcterms:description` foram adotados para apresentar o título do objeto representado pela string “Clavecin”@fr e a descrição do objeto fornecido pelo Cité de la Musique – Philharmonie De Paris. A Europeana determina o uso de um destes elementos de metadados como obrigatório e recomenda o uso de ambos, sempre que possível para ampliar as possibilidades de busca e recuperação do material.
- Para atribuir assuntos ao objeto, foram utilizadas informações do tesouro MIMO construído com a linguagem SKOS, representado pela propriedade `skos:prefLabel` que sinaliza o instrumento musical clavecin na língua francesa. O tesouro MIMO oferece uma taxonomia de assuntos referentes a instrumentos musicais, bem como o termo adotado em diversas línguas e o identificador de conceitos próximos ou correspondentes de outros esquemas conceituais. Estas informações estão ligadas ao objeto descrito na Europeana por meio do identificador 2251 inserido na própria URI do objeto.
- A localização do objeto real foi descrita por meio do metadado descritivo `dcterms:spatial` representado pela string “Florença”@it. Para representar o local foi adotada a classe `edm:Place` e para atribuir valor ao recurso foi realizada uma ligação entre o objeto e a fonte externa GeoNames representada pela propriedade `skos:prefLabel` que sinaliza o lugar associado ao objeto. Este dataset possibilita agregar ao objeto informações de dados geográficos como nomes de lugares em vários idiomas, elevação e população. O GeoNames está disponível na Web sob a licença Creative Commons.

Os recursos apresentados com o uso do esquema EDM permitem trazer mais informações para melhorar o acesso aos objetos originais. Segundo o EDM Primer (2013, p. 15), “eles podem permitir uma completa mudança de paradigma na forma como esses objetos são acessados, permitindo que o usuário navegue através de um espaço semântico de entidades contextuais antes de chegar aos objetos reais”.

A aplicação dos vocabulários consolidados às classes e propriedades do EDM é fundamental para viabilizar as representações e ligações entre as informações contidas em um determinado objeto do patrimônio cultural, bem como para garantir a evolução da Europeana de um esquema plano de metadados para um modelo de integração e enriquecimento de dados agregados em redes semânticas e a sua formalização como LOD (PATRICIO, 2012). Os dados da Europeana encontram-se publicados no LOD sob a licença Creative Commons Zero, os quais podem ser recuperados e acessados por meio da interface do tipo SPARQL EndPoin e baixados para reuso e agregações diversas nos formatos J-SON ou RDF.

5 Metodologia

Esta pesquisa caracteriza-se como qualitativa do tipo descritiva-documental baseada, principalmente, na família de documentos sobre o EDM, e busca descrever as relações semânticas inseridas nos elementos do modelo EDM, bem como destacar as vantagens de uso desse modelo para recuperação da informação na Web e, assim, incentivar a adoção de metodologias semânticas em projetos brasileiros. Para tanto, identificou os principais con-

ceitos e tecnologias da Web Semântica e Linked Data recomendados pelo W3C e analisou o memorial descritivo do modelo de dados da Europeana. O EDM foi detalhado em seus princípios e desenvolvimento, com ênfase nos elementos semânticos que modelam e apoiam a funcionalidade da Europeana.

O referencial teórico sobre Web Semântica, Linked Data e EDM foi construído a partir de artigos científicos recuperados nas bases de dados Scopus e Brapci, além de pesquisas em livros, artigos de revistas e sites da Web. A literatura sobre a Europeana e o padrão EDM foi consultada, principalmente, em sua família de documentos disponível no site da Europeana². O recorte temporal utilizado corresponde ao período de 2001 a 2017, no que se refere à Web Semântica e Linked Data. O referencial teórico sobre a Europeana e o EDM corresponde ao período de 2006 a 2018.

5 Conclusões

A análise das relações semânticas do padrão EDM evidenciou pontos de notório destaque no contexto da disponibilização de objetos culturais na Web. Dentre eles destaca-se a capacidade do EDM de agregar dados de instituições de diferentes domínios e disponibilizá-los em um único ambiente para recuperação, uso e reuso.

Outro ponto, pautado no modelo, é como a estrutura semântica do EDM se relaciona para possibilitar a agregação dos domínios culturais. O padrão EDM disponibiliza vocabulários próprios constituídos de classes e propriedades que dialogam com vocabulários amplamente aceitos. Essa harmonia permite a ligação entre diferentes instituições de forma a complementar e enriquecer as informações dos registros de um determinado objeto do patrimônio cultural. Essa ligação agrega conteúdos que possam existir mais em uma base de dados e não na outra, e vice-versa. O EDM permite usar URIs de fontes externas - como por exemplo tesouro MIMO, FOAF, VIAF, GeoNames dentre outros -, para atribuir valores omissos aos dados existentes e favorecer o processo de busca, resultando numa melhoria de qualidade da recuperação da informação. Sendo assim, pode-se afirmar que a estrutura proposta pelo EDM se ajusta aos princípios da Web Semântica e do Linked Data de enfatizar a reutilização e a ligação de recursos ricamente descritos na Web para enriquecer os dados de objetos de cultura, seja de bibliotecas, museus, arquivos e galerias de coleções audiovisuais.

Pesquisar sobre as relações semânticas aplicadas ao EDM, modelo responsável pela disponibilização do maior dataset de cultura no LOD, nos permitiu entender com maior riqueza de detalhes o papel de cada formato e como estes formatos podem estabelecer novas relações, capturando dados espalhados na Web, em diferentes contextos, para agregar informações aos seus recursos.

Embora o EDM esteja em desenvolvimento e as suas classes e propriedades estejam sendo implementadas de forma incremental, este modelo proporciona ao projeto Europeana um destaque inovador por disponibilizar dados abertos e conectados, conforme os princípios da Web Semântica e Linked Open Data.

Dessa maneira, pode-se afirmar que a Europeana, através do EDM, vem alcançando o objetivo de construir uma rica herança cultural europeia e torná-la acessível para o uso por pessoas e máquinas, em âmbito mundial, além de impulsionar o desenvolvimento do conceito das Humanidades Digitais, tema de grande interesse de investigação da área da Ciência da Informação, sobretudo das linhas de pesquisa que associam a informação e a tecnologia.

Espera-se que este estudo possa encorajar pesquisadores e idealizadores a fazerem uso de conceitos semânticos em projetos brasileiros, de forma que seus conjuntos de dados possam ser integrados ao LOD. A disponibilidade desses conjuntos de dados no LOD garante a interconexão cada vez maior de dados, aumenta a velocidade da circulação da informação entre os interessados e acelera novas descobertas.

Referências

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, New York, 17 may 2001.

BERNERS-LEE, T. Linked Data. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 13 jun. 2017.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data: the story so far. *International Journal on Semantic Web and Information Systems*, v.5, n.3, p. 1-22, 2009. Disponível em: <https://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 09 out. 2017.

CHARLES, V. Representing performing arts metadata in EDM. 2017. Disponível em: <<https://pro.europeana.eu/page/edm-for-performing-arts-metadata>>. Acesso em: 04 jan. 2019.

CONEGLIAN, C. S.; SANTAREM SEGUNDO, J. E. 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*, Florianópolis, v. 22, n. 48, p. 88-99, jan./abr. 2017. ISSN 1518-2924. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n48p88>. Acesso em: 09 out. 2017.

DOERR, M.; GRADMANN, S.; HENNICKE, S.; ISAAC, A.; MEGHINI, C.; SOMPEL, H. The Europeana Data Model (EDM). In: WORLD LIBRARY AND INFORMATION CONGRESS, 76., 2010, Gothenburg. Anais eletrônicos [...]. Gothenburg, Sweden: International Federation of Library Associations and Institutions, 2010. Disponível em: <https://www.ifla.org/past-wlic/2010/>. Acesso em: 21 dez. 2019.

EDM FACTSHEET. The Europeana data model for cultural heritage. [2013?]. Disponível em: <http://pro.europeana.eu/edm-documentation>. Acesso em: 09 out. 2017.

EDM MAPPING GUIDELINES. 2016. Disponível em: <http://pro.europeana.eu/edm-documentation>. Acesso em: 09 out. 2017.

EDM PRIMER. Europeana Data Model Primer. 2013. Disponível em: <http://pro.europeana.eu/edm-documentation>. Acesso em: 09 out. 2017. Editado por Antoine Isaac.

EUROPEANA PRO. Europeana Data Model. Disponível em: <<https://pro.europeana.eu/resources/standardization-tools/edm-documentation>>. Acesso em: 4 jan. 2018.

JEFCOATE, Graham. Gabriel : gateway to Europe's national libraries. *Emerald Insight*, v.40, n.4, p. 325-333. 2006. Disponível em : <<https://doi.org/10.1108/00330330610707908>>. Acesso em: 4 jan. 2019.

ISOTANI, S.; BITTENCOURT, I. I. Dados abertos conectados. São Paulo: Novatec, 2015.

OPEN KNOWLEDGE INTERNATIONAL. The open data handbook. 2004. Disponível em: <<http://opendatahandbook.org/guide/en/what-is-open-data/>>. Acesso em: 24 set. 2017.

PATRÍCIO, H. S. A Europeana e a agregação de metadados na web: análise dos esquemas ESE/EDM e da aplicação de standards da web semântica a dados de bibliotecas. In: *Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*, 11., 2012, Lisboa. Anais... Lisboa, 2012.

RYAN, C.; GRANT, R.; CARRÁIAN, E.; COLLINS, S.; STEFAN, D.; LOPES, N. Linked data authority records for Irish place names. In *J Digit Libr, Berlin Heidelberg*, v. 15, p. 73-85, abr. 2015.

SANTAREM SEGUNDO, J. E. Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas internacionais. In: *ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 16., 2015, João Pessoa. Anais eletrônicos... João Pessoa: UFPB, 2015. p. 219-239.

W3C. World Wide Web Consortium. Disponível em: <<https://www.w3.org/>>. Acesso em: 04 jan. 2019.

W3C. SPARQL: 1.1 visão geral. 2013. Disponível em: <<https://www.w3.org/TR/sparql11-overview/>> . Acesso em: 24 jun. 2018.

W3C. OWL Web Ontology Language. W3C recommendation. 2004. Disponível em: <<https://www.w3.org/TR/owl-ref/>>. Acesso em: 08 out. 2018.

WINER, D.; ROCHA, I.E. Europeana: um projeto de digitalização e democratização do patrimônio europeu. *Patrimônio e Memória*, São Paulo: UNESP, v.9, n.1, p.113-127, jan./jun. 2013. ISSN 1808-1967.

Notas al final

1 i2010 é o um quadro estratégico da Comissão Europeia que define as grandes orientações políticas para a sociedade da informação e os media. Com a iniciativa i2010, a Comissão aborda de modo integrado a sociedade da informação e as políticas audiovisuais na União Europeia. Tem o objetivo de coordenar a ação dos Estados-Membros para facilitar a convergência digital e responder aos desafios ligados à sociedade da informação. Disponível em: <http://eur-lex.europa.eu>

2 Documentação do Modelo de Dados da Europeana: <http://pro.europeana.eu/edm-documentation>

Dados dos autores

Luciana Candida Silva

Doutora em Ciência da Informação pela Universidade Estadual Paulista Júlio Mesquita Filho (UNESP). Mestre em Ciência da Informação pela Universidade de Brasília (UNB) e graduada em Biblioteconomia pela Universidade Federal de Goiás (UFG).

candida.luciana@gmail.com

José Eduardo Santarem Segundo

Professor Doutor. Departamento de Educação, Informação e Comunicação. USP - FFCLRP. Ribeirão Preto, SP.

santarem@usp.br

Nuno Freire

Investigador no Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC-ID). Doutorado em Engenharia Informática e de Computadores pelo Instituto Superior Técnico da Universidade de Lisboa.

nuno.freire@tecnico.ulisboa.pt

Received: 2019-03-19

Accepted: 2021-07-31



This work is licensed under a Creative Commons Attribution 4.0 United States License.



This journal is published by the University Library System of the University of Pittsburgh as part of its D-Scribe Digital Publishing Program and is cosponsored by the University of Pittsburgh Press