

AVALIAÇÃO DO EUROPEANA DATA MODEL PARA O ENRIQUECIMENTO SEMÂNTICO DE OBJETOS

EVALUATION OF THE EUROPEANA DATA MODEL FOR THE SEMANTIC ENRICHMENT OF OBJECTS

Claudiane Emanuele Nazário 
Universidade Federal de Minas Gerais

Célia da Consolação Dias 
Universidade Federal de Minas Gerais

RESUMO

A Europeana desenvolveu o modelo de dados Europeana Data Model (EDM), visando garantir a interoperabilidade entre dados de diferentes fornecedores, utilizando os princípios do Linked Data. Este trabalho tem por objetivo avaliar as contribuições do EDM para o enriquecimento semântico de objetos publicados na WEB por meio do Linked Data. Para atingir este objetivo foi criada uma Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos (TRESO), desenvolvida durante a pesquisa de mestrado em Gestão e Organização do Conhecimento. A Matriz TRESO é composta por sete critérios que analisam como o modelo avalia as técnicas e os recursos de enriquecimento semântico de objetos. Como resultado é demonstrado o grau de atendimento do EDM a cada um dos critérios da Matriz TRESO e indicadas às recomendações de melhoria para que o modelo possa aprimorar o processo de enriquecimento semântico de objetos publicados em Linked Data.

Palavras-chave: EDM, Europeana Data Model, Enriquecimento Semântico, Linked Data.

ABSTRACT

Europeana has developed the Europeana Data Model (EDM) data model to ensure interoperability between data from different vendors using Linked Data principles. This work aims to evaluate the contributions of EDM to the semantic enrichment of objects published on the WEB through Linked Data. To achieve this goal, a Matrix of Techniques and Resources for the Semantic Enrichment of Objects (TRESO), developed during the masters research in Knowledge Management and Organization, was created. The TRESO Matrix is composed by seven criteria that analyze how the model evaluates the techniques and resources of semantic enrichment of objects. As a result it is demonstrated the degree of EDM service to each of the criteria of the TRESO Matrix and indicated to the improvement recommendations so that the model can improve the process of semantic enrichment of objects published in Linked Data.

Keywords: EDM, Europeana Data Model, Semantic Enrichment, Linked Data.

1. INTRODUÇÃO

A Europeia desenvolveu o modelo de dados *Europeana Data Model* (EDM), visando garantir a interoperabilidade entre dados de diferentes fornecedores provenientes de diversas comunidades, que utilizam formatos de metadados distintos para descrever recursos de patrimônio cultural. (CHARLES; ISAAC; FREIRE, 2014)

De acordo com HASLHOFER e ISAAC (2011) o EDM utiliza uma estrutura baseada na Web Semântica que permite o enriquecimento dos dados a partir de fontes externas, utilizando o *Linked Data*.

A adoção do *Linked Data*, em conjunto com técnicas de enriquecimento semântico favorecem a qualidade de dados disponibilizados, bem como facilitam a interoperabilidade de dados entre diferentes fontes.

Entender como funcionam os modelos de dados, os recursos por eles utilizados, as tecnologias empregadas e as dificuldades enfrentadas por esses modelos para realizar o enriquecimento semântico para sua publicação em *Linked Data*, se constituem de um importante objeto de estudo para a Biblioteconomia e Ciência da Informação (BCI).

Este artigo apresenta parte dos resultados de uma pesquisa de mestrado em Ciência da Informação, que buscou desenvolver uma proposta de metodologia para avaliar o enriquecimento semântico de objetos publicados na Web através do *Linked Data*. Neste trabalho são retratadas, especificamente, as análises realizadas no EDM utilizando a proposta referenciada e, tem por objetivo geral avaliar as contribuições desse modelo de dados para o enriquecimento semântico de objetos publicados na Web através do *Linked Data*. Para atingir esse objetivo foi aplicada a Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos (Matriz TRESO), desenvolvida durante a pesquisa de mestrado em Gestão & Organização do Conhecimento na Escola de Ciência da Informação da Universidade Federal de Minas Gerais, Brasil.

Os objetivos específicos desse trabalho são: a) estudar os aspectos que favorecem o enriquecimento semântico de dados, utilizados pelo EDM e b) analisar a Matriz Treso como mecanismo para avaliação de modelos de dados.

1.1 Problema de Pesquisa e Justificativa

A literatura apresenta diferentes modelos de dados que realizam o enriquecimento semântico de objetos, visando sua publicação no *Linked Data*. No entanto, a forma com que esse processo de enriquecimento é realizado e de que maneira ele contribui para a qualificação desses dados ainda carece de maiores discussões no âmbito da BCI.

A questão investigada nesse estudo foi avaliar quais as contribuições do EDM para o enriquecimento semântico de objetos publicados na Web através do *Linked Data*? A resposta a esse questionamento teve como foco contribuir tanto para a identificação de melhorias no próprio EDM, no que se refere à semântica e qualidade dos dados publicados na web, quanto para testar a Matriz TRESO como um instrumento de avaliação das técnicas e recursos para o enriquecimento semântico de objetos.

Ampliar as discussões sobre semântica, qualidade e interoperabilidade de dados favorecem não só a análise de modelos de dados existentes, mas também a auxilia na construção de novos modelos voltados para a publicação de dados em *Linked Data* e enriquecidos semanticamente.

2. REFERENCIAL TEÓRICO

A seguir apresenta-se um panorama teórico-contextual acerca dos conceitos e temas que fundamentaram essa pesquisa. Expõem-se as ideias de diferentes autores que contribuíram com os insumos necessários para uma discussão qualitativa para o estudo realizado.

Inicia-se esse capítulo com as contribuições do campo de estudo do enriquecimento semântico. A seguir são apresentados conceitos para a publicação de dados em *Linked Data*. Finalizando o capítulo apresentam-se o modelo de dados EDM utilizado para testar os objetivos dessa pesquisa.

2.1 Enriquecimento Semântico

O enriquecimento semântico pode ser compreendido como um processo de atribuição de maior significado aos dados e metadados, através do uso da semântica atribuída por vocabulários pré-existentes, sinônimos e informações de proveniência. Esse processo tem por objetivo facilitar a compreensão, a integração e o processamento dos dados por pessoas e máquinas. (Lira, 2014)

Segundo Torino et. al (2020), para garantir a compreensão de homens e máquinas no processo de disponibilização de dados abertos utilizando o *Linked Data* é necessário o planejamento desse processo e o enriquecimento semântico desses dados. A medida em que a web semântica se desenvolve, novos conceitos, técnicas e tecnologias são incorporados com o objetivo do tratamento semântico dos dados disponibilizados na web.

2.2 Linked Data

Em 2006, Tim Bernes-Lee recomendou um conjunto de princípios e boas práticas para publicação de dados denominado *Linked Data*, visando suprir as necessidades advindas da evolução da web, tais como manipular um grande volume de dados e facilitar o acesso à informação. O *Linked Data* tem como objetivo fundamental facilitar a integração de dados de diferentes fontes, de forma a torná-los compreensíveis também para as máquinas (BIZER; HEATH; BERNERS-LEE, 2009) e podem ser resumidos da seguinte forma.

1. Os recursos na web devem ser denominados, utilizando URIs (*Uniform Resource Identifier*);
2. Para que os usuários possam localizar esses recursos devem ser utilizadas URIs HTTP (*Hypertext Transfer Protocol*);
3. Devem ser fornecidas informações úteis, utilizando os padrões RDF (*Resource Description Framework*), SPARQL (serviços de consulta que utilizam o protocolo HTTP) quando o usuário procurar uma URI;
4. Devem ser incluídos links para outras URIs, possibilitando a descoberta de novos recursos na web. (BERNERS-LEE, 2006, p.1)

Várias empresas e instituições utilizam o *Linked Data* para a publicação de seus dados na web semântica uma delas é a Europeia que desenvolveu um modelo de dados para disponibilizar dados de diferentes provedores.

2.3 Europeana Data Model

A Europeia é uma plataforma digital que reúne e estrutura dados de diferentes instituições de patrimônios culturais. Sua função inclui tanto a disponibilização de todo seu conteúdo de modo estruturado por meio da web, quanto de mecanismos para a reutilização e acesso aberto

desses dados através de seu modelo de dados, o *Europeana Data Model* (EDM), de modo a facilitar o uso de recursos digitalizados do patrimônio cultural. (TRIQUES; ARAKAKI, 2021).

O EDM tem por objetivo padronizar a descrição dos recursos do patrimônio cultural de instituições, como bibliotecas, arquivos, museus e galerias utilizando o modelo *Linked Open Data* (LOD). (ARAKAKI, 2016). De acordo com a Europeana (2011), os requisitos que guiaram a construção do modelo EDM, são:

- Distinção entre os objetos e suas representações digitais, ou seja, o objeto real seja uma pintura, escultura, livro, etc. possui identificação diferente de sua representação na web;
- Distinção do objeto e seu registro de metadados, ou seja, o objeto real é diferente dos dados cadastrados pelos provedores de conteúdo;
- Possibilidade de vários registros para um mesmo objeto, ainda que contenham declarações divergentes ou contraditórias, ou seja, os registros do metadados feitos por diferentes provedores podem estar disponíveis para o usuário, ainda que apresentem descrições ou conceitos opostos;
- Suporte para objetos que são compostos de outros objetos, permitindo que esses sejam recuperados em conjunto, através de suas relações semânticas ou individualmente;
- Suporte para recursos contextuais, incluindo os conceitos de vocabulários controlados, permitindo a reutilização de termos e relações.

As premissas utilizadas pela Europeana para o desenvolvimento do EDM consideram não só técnicas de enriquecimento semântico de objetos, mas também a disponibilização desses dados de forma aberta. Em virtude dessas premissas o EDM foi utilizado na parte empírica deste estudo para avaliar como o processo de enriquecimento semântico é realizado por esse modelo de dados para a publicação em *Linked Data*.

3. METODOLOGIA

Para realização desse estudo foi realizada uma pesquisa aplicada e exploratória utilizando os seguintes procedimentos técnicos:

- Pesquisa bibliográfica e documental para identificar os principais requisitos e técnicas de enriquecimento semântico de objetos para publicação de dados em *Linked Data* na literatura.

- Construção da Matriz TRESO, utilizando tanto materiais bibliográficos, quanto pesquisa direta no *dataset* da Europeana disponível no LOD, e no catálogo da *Library of Congress* existente na Web, sendo que a pesquisa foi realizada diretamente em fontes primárias de informação, conforme descrito no item 3.1 desse artigo a seguir.
- Aplicação da Matriz TRESO no modelo EDM, visando identificar quais os recursos e técnicas de enriquecimento semântico foram utilizados no modelo.

Pontuação do modelo EDM em cada critério da matriz, atribuindo notas de 0 a 4, conforme apresentado no item 4 desse artigo.

3.1 A Matriz TRESO

A Matriz TRESO foi construída em 2018 com a finalidade de permitir a verificação de quais os recursos e técnicas de enriquecimento semântico estavam presentes no modelo de dados desenvolvidos para publicação de objetos em *Linked Data*. Essa matriz foi desenvolvida a partir de elementos identificados na literatura, considerando os estudos de (LIRA, 2014); (SILVA, 2016); (Silva, 2014); (PITA; PAIXÃO, 2010); e (SORRENTINO et al., 2013).

O quadro 1 a seguir apresenta a Matriz criada contendo os sete critérios para avaliação dos recursos e técnicas de enriquecimento semântico, bem como os parâmetros definidos para estabelecer a pontuação para cada um dos critérios.

Quadro 1 - Matriz TRESO

MATRIZ DE TÉCNICAS E RECURSOS PARA O ENRIQUECIMENTO SEMÂNTICO DE OBJETOS					
CRITÉRIO DE AVALIAÇÃO	INEXISTENTE	EM DESENVOLVIMENTO	BAIXO	ATENDE	ATENDIMENTO TOTAL
Pontuação	0	1	2	3	4
1. Utilização de recursos de anotação semântica para o enriquecimento de objetos publicados em <i>Linked Data</i>	O modelo de dados não permite o uso de anotação semânticas de qualquer natureza.	O modelo de dados permite a inclusão de anotações por meio de <i>tags</i> e textos livres sem nenhum controle terminológico e passível de múltiplas interpretações.	O modelo de dados permite a inclusão de anotações por meio de <i>propriedades</i> , não permitindo muitos relacionamentos entre os recursos.	O modelo de dados realiza a anotação semântica utilizando vocabulários controlados, permitindo a conexão com dados de diferentes fontes que utilizem o mesmo vocabulário.	O modelo de dados realiza a anotação semântica utilizando ontologias, de modo a identificar os objetos e suas relações favorecendo a conexão de dados de diferentes fontes, permitindo seu enriquecimento para publicação em <i>Linked Data</i> , vocabulário.
2. Reuso de metadados para facilitar as atividades de publicação de dados em <i>Linked Data</i> , otimizando o trabalho do publicador	O modelo de dados não permite a reutilização de metadados de outras fontes ou vocabulários.	O modelo de dados permite a reutilização de metadados em uma classe específica ou agrega dados e metadados de uma única fonte.	O modelo de dados permite a reutilização de metadados em até duas classes específica ou realiza a agregação de dados e metadados de mais uma fonte.	O modelo de dados permite a reutilização de diferentes metadados em todas as classes e realiza a agregação de dados e metadados de diferentes fontes e vocabulários.	O modelo de dados permite a reutilização de diferentes metadados em todas as classes e propriedades do modelo, bem como realiza a agregação de dados e metadados de diferentes fontes e vocabulários e ontologias.
3. Links entre as combinações semânticas dos dados e metadados com outros recursos da web.	Inexistência de links entre os dados e metadados com outros recursos da web.	Existência de links entre as combinações semânticas dos dados e metadados com um <i>dataset</i> específico.	Existência de links entre as combinações semânticas dos dados e metadados com até 2 <i>datasets</i> específico.	O modelo prevê a integração entre as combinações semânticas dos dados e metadados com quaisquer <i>datasets</i> da web de dados.	O modelo prevê a integração entre as combinações semânticas dos dados e metadados, utilizando ontologias para conectar com quaisquer <i>datasets</i> da web de dados publicados no projeto LOD2
4. Modelagem de dados num formato semântico estruturado, permitindo sua manipulação por aplicações que consomem esses modelos, ampliando as possibilidades de conexão com outros <i>Datasets</i> do LOD.	O modelo de dados não prevê a utilização de formato semântico estruturado, impossibilitando sua manipulação por agentes de software.	Em apenas uma classe do modelo são utilizados formatos semânticos e estruturados como o RDF/XML.	Em apenas até duas classe do modelo são utilizados formatos semânticos e estruturados como o RDF/XML.	Os dados são modelados num formato semântico estruturado, como o RDF/XML, beneficiando sua manipulação por aplicações que consomem esses modelos de dados	Os dados são modelados num formato semântico estruturado, como o RDFS e OWL beneficiando sua manipulação por aplicações que consomem esses modelos de dados

MATRIZ DE TÉCNICAS E RECURSOS PARA O ENRIQUECIMENTO SEMÂNTICO DE OBJETOS					
CRITÉRIO DE AVALIAÇÃO	INEXISTENTE	EM DESENVOLVIMENTO	BAIXO	ATENDE	ATENDIMENTO TOTAL
Pontuação	0	1	2	3	4
5. Utilização de ferramentas para o enriquecimento semântico.	O modelo realiza o enriquecimento semântico de forma manual.	O enriquecimento semântico é realizado de forma semiautomática apenas pelos criadores do modelo	O enriquecimento semântico é realizado de forma semiautomática por publicadores e usuários.	Todo o processo de enriquecimento semântico é feito de forma automática, utilizando ferramentas proprietárias para realizar anotações ou vinculação de dados.	Todo o processo de enriquecimento semântico é feito de forma automática, utilizando ferramentas livres para realizar anotações ou vinculação de dados.
6. Utilização de interface gráfica para apoiar a execução do processo de enriquecimento semântico e publicação de dados.	Inexistência de interface gráfica para facilitar o enriquecimento semântico pelo publicador.	A interface gráfica é utilizada para uma única função, sendo o processo de enriquecimento semântico semiautomático.	A interface gráfica é utilizada para algumas funções, sendo o processo de enriquecimento semântico semiautomático.	Os publicadores de dados utilizam a interface gráfica para apoiar a execução do processo de enriquecimento semântico.	Os publicadores de dados e usuários da informação utilizam a interface gráfica para apoiar a execução do processo de enriquecimento semântico.
7. Relações de sinonímia (equivalência), associação e hierarquia entre o metadado e o termo correspondente em outros vocabulários utilizados.	O sistema não apresenta relações de sinonímia (equivalência), associação e hierarquia entre o metadado e o termo correspondente em outros vocabulários utilizados.	O modelo estabelece relações de sinonímia (equivalência), hierarquia, associação com alguns termos, mas em uma única classe.	O modelo estabelece relações de sinonímia (equivalência) com todos os termos do modelo.	Utilização de relações de sinonímia (equivalência) entre o metadado e o termo correspondente com outros vocabulários utilizados, reduzindo a ambiguidade.	Utilização de relações de sinonímia (equivalência), associação e hierarquia entre o metadado e o termo correspondente em outros vocabulários utilizados.

Fonte: Elaborado pelas autoras

4. APLICAÇÃO DA MATRIZ TRESO NO EDM E AVALIAÇÃO

Nesta seção, foi identificado o modo de atendimento do EDM a cada um dos critérios listados na Matriz TRESO (Quadro 1) para avaliação dos recursos e técnicas para o enriquecimento semântico de objetos. Ressalta-se que esse artigo retrata a avaliação realizada durante a pesquisa de mestrado no ano de 2018.

3.2 Critério 1 - Anotação semântica

O primeiro critério da Matriz TRESO em que o EDM foi avaliado, se refere ao modo com que o modelo utiliza o recurso de anotação semântica, ou seja, o modo como adiciona as informações semânticas para atribuir significado aos recursos publicados na Web.

Após a análise observou-se que o EDM realiza a anotação tanto por meio de tags (literais), quanto através de referências a vocabulários externos, utilizando o RDF. Esta forma de anotação possibilita ao modelo interligar vários recursos através de links (relações), permitindo a navegação de um recurso para outro através destes links pelos usuários.

De acordo com (CONEGLIAN; SANTARÉM, 2017), o processo de anotação semântica é realizado pelo escritório da Europeia. As informações relevantes dos objetos são relacionadas a tesouros e *datasets*, permitindo a ligação desses objetos com as *URIs* existentes, de modo a conectar os dados com outras bases existentes.

Como não foi possível observar se a forma de anotação adotada pelo EDM a permite utilização de ontologias, que possibilitem aos sistemas a realização de inferências para localizar outros recursos, não foi atribuída a pontuação máxima neste item, sendo, portanto, atribuída a **nota 3 (três) - atendimento satisfatório**.

3.3 Critério 2 - Reuso de metadados

Este critério visa avaliar a reutilização de metadados de outras fontes para facilitar a publicação de dados em Linked Data para otimizar o trabalho do publicador.

Uma das premissas do EDM é reutilizar termos e relações de diferentes vocabulários, por meio de referências, ou seja, o modelo utilizar as propriedades em seus vocabulários de origem. Deste modo, o EDM faz menção a elementos utilizados em outros vocabulários como Dublin Core, Open Archives

Object Reuse and Exchange Model (OAI-ORE) e Simple knowledge Organization System – SKOS dentre outros.

Em todas as classes do modelo é realizado o reaproveitamento de metadados de outros vocabulários. Das 116 propriedades, 85 são reutilizadas de outros padrões, totalizando 73% das propriedades do modelo, conforme apontado no Quadro 2 seguir:

Quadro 2 – Reaproveitamento de metadados pelo EDM

Classe	Total de propriedades	Propriedades reaproveitadas	% de reaproveitamento
Provided Cultural Heritage Object (Objeto Patrimônio Cultural)	50 propriedades	37 Dublin Core 1 OWL	76%
Web Resource (Recursos da web)	15 propriedades	12 Dublin Core 1 OWL	86%
Aggregation (Agregação)	10 propriedades	1 Dublin Core	10%
Agent (agente)	22 propriedades	3 Skos 4 Dublin Core 9 RDA 1 Foaf 1 OWL	81%
Place (local)	10 propriedades	3 Geonames (wgs) 3 SKOS 2 Dublin Core 1 OWL	90%
Timespan ou Period (período)	9 propriedades	3 Skos 2 Dublin Core 1 OWL	66%

Fonte: Elaborado pelas autoras a partir da coleta de dados

A opção do EDM em reutilizar propriedades de vocabulários existentes, consolidados e disponíveis no LOD, ao invés de replicá-los, permite maior facilidade para integração e consequente publicação de objetos, tendo em vista que vários *datasets* utilizam alguns desses vocabulários.

Deste modo, como o EDM reutiliza metadados em todas as classes de modelo foi atribuída a **nota 4 (quatro), Atendimento Total ao critério.**

3.4 Critério 3 - Links entre dados e metadados do modelo com outros recursos da web

Este critério avalia a existência de links entre as combinações semânticas dos dados e metadados do EDM com outros recursos da web. Neste critério, a quantidade de conexões com outros datasets está

diretamente relacionada à facilidade para publicar os dados em Linked Data e realizar a interoperabilidade de dados.

O EDM realiza a conexão com vários outros vocabulários disponíveis na Web Dados. Segundo o Linked Open Vocabulary (LOV), o modelo apresenta treze destes links de saída, conforme apresentado no Quadro 3 a seguir:

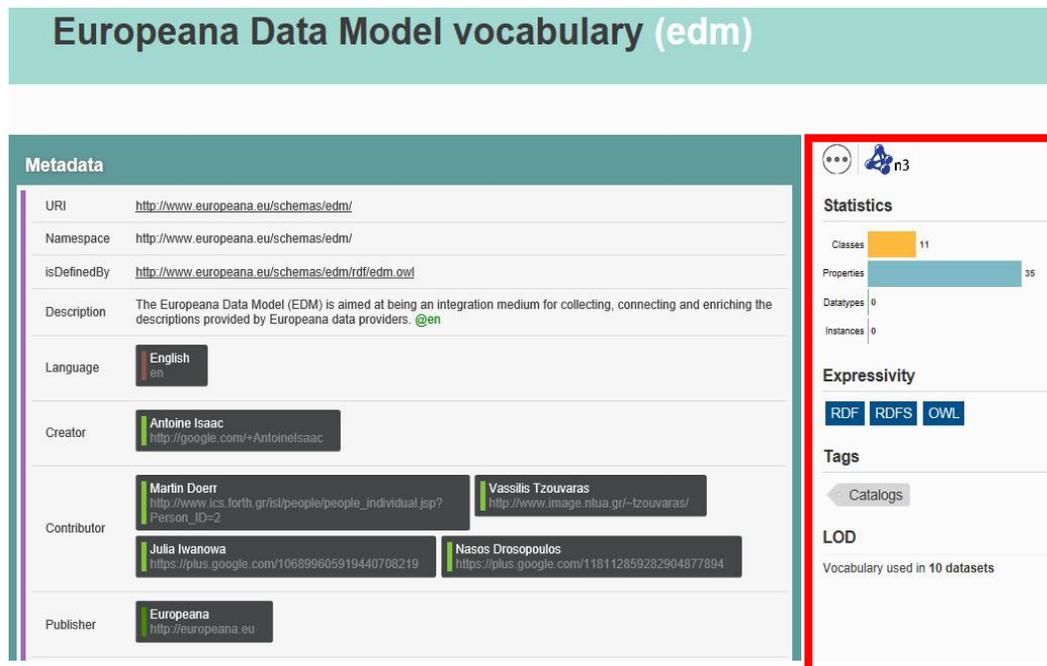
Quadro 3 – Características do EDM apresentadas no *Linked Open Vocabulary* (LOV)

Critério	EDM
Utilização do Vocabulário no LOD	10 datasets
Links de saída	13 links
Links de entrada	1 link
Classes utilizadas	11 classes
Propriedades utilizadas	35 propriedades

Fonte: Open Knowledge Fundation, 2018

A Figura 1 - apresenta as estatísticas de utilização do modelo no contexto da web semântica e do *Linked Open Data* publicados no site do LOV.

Figura 1- Utilização do EDM no *Linked Open Vocabulary* (LOV)



Fonte: Open Knowledge Fundation, 2018

O EDM apresenta link de entrada com o vocabulário a Simple Service Status Ontology (SSSO), uma ontologia RDF baseada em eventos. Essa ontologia apresenta relacionamento com outras ontologias tais como a CIDOC-CRM (*CIDOC Conceptual Reference Model*) e DOLCE+DnS *Ultralite ontology* (DUL).

Em relação a este quesito, foi possível observar que o EDM atende totalmente ao critério estabelecido, por dois motivos: 1) porque o modelo realiza conexões com vários *datasets* do *Linked Open Data* (LOD) 2) porque ele prevê a integração entre as combinações semânticas dos dados e metadados, utilizando ontologias para conectar com quaisquer *datasets* publicados no LOD. Deste modo foi atribuída a **nota 4 (quatro) - Atendimento Total ao critério.**

3.5 Critério 4 - Modelagem de dados num formato semântico estruturado

No que se refere ao critério Modelagem de dados num formato semântico estruturado, o EDM é modelado de modo a transformar seus metadados em URI, representados no padrão RDF/XML. O enriquecimento dos dados e a associação destes com os objetos culturais, são realizados para a conversão e a integração de dados de fontes existentes no LOD.

O EDM também utiliza a linguagem OWL e os conceitos da ontologia CIDOC CRM no processo de modelagem. A utilização destas ontologias resulta em uma maior qualidade dos dados e dos relacionamentos entre os mesmos.

Segundo ISOTANI e BITTENCOURT (2015) as relações semânticas criadas pelas ontologias podem ser utilizadas pelo *Linked Data* para ampliar as conexões entre os recursos na web.

Conclui-se que no critério de modelagem de dados em um formato estruturado, a utilização de ontologias foi o diferencial adotado pelo EDM que resultou em uma pontuação total, sendo atribuída **nota 4 (quatro) - Atendimento Total** ao critério, porque os dados são modelados num formato semântico estruturado, como OWL, bem como utiliza conceitos do CIDOC CRM, beneficiam sua manipulação por aplicações que consomem esses modelos de dados.

3.6 Critério 5 - Utilização de ferramentas para o enriquecimento semântico

A Europeana desenvolveu uma ferramenta tecnológica de enriquecimento, cujo objetivo é realizar a vinculação de dados de modo automático. São realizadas referências entre os objetos do EDM com as URIs de vocabulário externos, assim como são recuperados dados multilíngues e semânticos relacionados a um determinado conceito do vocabulário. Se um vocabulário é representado pelo

SKOS, que também é reutilizado no EDM para descrever dados conceituais, todo o processo de enriquecimento é facilitado.

Segundo CHARLES, ISAAC e FREIRE (2014) a Europeia também realiza o enriquecimento automático com outros vocabulários e datasets como GEMET, GeoNames e DBpedia criando links para objetos na Europeia.

Apesar do EDM possuir ferramentas para apoiar no processo de enriquecimento semântico de forma automática, o fato das ferramentas serem proprietárias dificulta sua utilização pelos usuários. Deste modo, nesse item foi atribuída **nota 3 - atendimento satisfatório** - uma vez que as anotações ou vinculação de dados são realizadas por meio de ferramentas proprietárias e não utilizando software livre.

3.7 Critério 6 - Utilização de interface gráfica

Esse critério considera a utilização de interface gráfica para processo de enriquecimento semântico e publicação de dados.

(UREN et. al., 2005) afirmam que para realizar um processo colaborativo é necessário fornecer aos publicadores, interfaces gráficas que simplifiquem o processo de anotação. Ainda segundo os autores a disponibilização de uma interface de anotação integrada com o ambiente de criação, leitura, compartilhamento e edição de informações poderia facilitar o processo de enriquecimento semântico.

Tendo em vista que não foi possível identificar a utilização de interface gráfica para o enriquecimento semântico de objetos no modelo EDM, foi atribuída **nota zero - não atende ao critério**.

3.8 Critério 7 – Relacionamento entre os metadados do modelo e termos de outros vocabulários.

Este critério se refere à utilização de relações de equivalência, associação e hierarquia entre o metadado e o termo correspondente em outros vocabulários utilizados.

No EDM as conexões entre os vocabulários são realizadas por meio do RDF, no qual é possível identificar os tipos de relacionamentos existentes entre os objetos, através de propriedades relacionadas no quadro 4 a seguir:

Quadro 4 – Tipos de Relacionamentos existente no Modelo EDM

Tipo de Relacionamento	Propriedade correspondente
Hierarquia	dcterms:hasPart, dcterms:isPartOf, edm:hasType edm:type
Associação	edm:isSimilarTo
Sinonímia ou equivalência	owl:sameAs

Fonte: Elaborado pelas autoras a partir da estrutura do EDM

Neste item foi atribuído ao EDM **nota 4 (quatro) - Atendimento Total** ao critério, porque as propriedades utilizadas apresentam relações equivalência, associação e hierarquia entre o metadado do modelo e o termo correspondente em outros vocabulários que poderão ser conectados a partir do LOD.

5. ANÁLISE DE RESULTADOS

Após a análise e aplicação da Matriz TRESO no EDM foi possível observar que o comportamento do modelo em cada um dos critérios propostos, bem como suas contribuições para o processo de enriquecimento semântico listados a Seguir:

- Anotação Semântica - O EDM mostrou atendimento satisfatório a esse critério uma vez que utiliza links para acesso a outros vocabulários, no entanto não foi possível comprovar o uso de ontologias. Segundo Silva (2014) as ontologias são mais vantajosas do que outros tipos de anotação, na medida em que permitem relações mais explícitas e significativas entre diferentes tipos de dados, favorecendo sua interpretação por agentes de software.
- A reutilização de metadados de outros modelos em seu contexto original - Critério em que o EDM demonstrou excelência, em virtude de o modelo fazer o reuso de propriedades de outros vocabulários em todas as suas classes. De acordo com GRUBER (1995) a utilização de um padrão de vocabulário comum para troca de dados permite a reutilização do conhecimento e facilita a interoperabilidade entre diferentes sistemas. Outros modelos podem ser conectados ao EDM e fazer uso desses metadados em seu contexto original criando novos links semânticos.

- A conexão com outros recursos da web – Neste item a quantidade de links do EDM apresentada no LOV foi determinante para a pontuação obtida pelo modelo, na medida em que o portal demonstra uma ampla utilização do modelo no LOD. Trata-se de uma boa prática adotada pelo EDM que pode ser replicada por outros modelos.
- A modelagem de dados em um formato semântico estruturado, permitindo sua compreensão por homens e máquinas e sua consequente reutilização – O uso de ontologias realizado pelo EDM, lhe conferiu pontuação total em virtude do importante papel desempenhado pelas ontologias no processo de enriquecimento semântico, tanto para facilitar a interpretação dos dados por homens e máquinas quanto para realizar inferências e ampliar as possibilidades de pesquisa.
- A explicitação dos relacionamentos entre os metadados do modelo e termos de outros vocabulários – Neste critério, como o EDM possui algumas propriedades que indicam claramente o tipo de relação entre os objetos e suas propriedades o mesmo obteve pontuação total. Essa característica também pode ser replicada por outros modelos de dados.
- Utilização de ferramentas para facilitar o trabalho do publicador e permitir a manipulação de grandes volumes de dados – Apesar do EDM possuir ferramentas que facilitam o processo de publicação o fato das mesmas serem proprietárias dificulta sua utilização por outros usuários e instituições. A utilização de Software Livre pelo EDM facilitaria a replicação do processo de publicação adotado pelo modelo para outras instituições de menor poder aquisitivo.
- Existência de interface gráfica para facilitar o trabalho do publicador. Critério este que não pode ser comprovado durante a avaliação do modelo. A utilização de uma interface gráfica permitiria maior velocidade e facilidade na publicação dos dados pelos usuários.

Apresenta-se a seguir o Quadro 5 contendo uma síntese dos resultados e da pontuação obtida pelo EDM em cada um dos critérios propostos pela Matriz TRESO.

Quadro 5 – Pontuação obtida pelo EDM nos critérios da Matriz Treso.

Critério de Avaliação	Pontuação
Utilização de recursos de anotação semântica para o enriquecimento de objetos publicados em <i>Linked Data</i>	3
Reuso de metadados para facilitar as atividades de publicação de dados em <i>Linked Data</i> , otimizando o trabalho do publicador	4
Links entre as combinações semânticas dos dados e metadados com outros recursos da web.	4
Modelagem de dados num formato semântico estruturado, permitindo sua manipulação por aplicações que consomem esses modelos de dados, ampliando as possibilidades de conexão com	4

Critério de Avaliação	Pontuação
outros <i>datasets</i> do LOD.	
Utilização de ferramentas para o enriquecimento semântico.	3
Utilização de interface gráfica para apoiar a execução do processo de enriquecimento semântico e publicação de dados.	0
Relações de sinonímia (equivalência), associação e hierarquia entre o metadado e o termo correspondente em outros vocabulários utilizados.	4
RESULTADO	22

Fonte: elaborado pelas autoras

Dos 28 pontos previstos o EDM atingiu 22 pontos o que indica um alto nível de atendimento aos critérios propostos pela Matriz TRESO , no que se refere à aplicação de técnicas e recursos para o enriquecimento semântico de objetos para sua publicação em *Linked Data*, contudo ainda existem pontos que podem implementados para a melhoria do processo de enriquecimento semântico adotado modelo.

6. CONSIDERAÇÕES FINAIS

Esse artigo teve como propósito avaliar quais as técnicas e recursos para o enriquecimento semântico utilizados pelo EDM para a publicação de dados em *Linked Data*. Após a análise do modelo por meio da aplicação da Matriz TRESO foi possível afirmar que o modelo EDM possui a maior parte dos elementos necessários para o enriquecimento semântico de objetos para sua publicação em *Linked Data* propostos pela matriz. Contudo, o modelo ainda pode incorporar melhorias, através da utilização de interface gráfica para agilizar o trabalho de publicação de grandes volumes de dados e utilizar ontologias no processo de enriquecimento semântico.

As discussões acerca do comportamento do modelo em cada um dos critérios elencados na Matriz TRESO, também permitiu a validação desse instrumento como uma ferramenta que pode ser utilizada para avaliar técnicas e recursos de enriquecimento semântico adotados por outros modelos, bem como orientar a construção de novos modelos que tenham o objetivo de publicar dados com qualidade no ambiente da web, utilizando o *Linked Data*.

Acredita-se que os resultados obtidos foram satisfatórios e tornam viáveis a continuidade da pesquisa iniciada no âmbito da BCI, na medida em que novos modelos podem ser desenvolvidos ou

aprimorados com o presente trabalho, bem como o modelo analisado pode incorporar as melhorias propostas nesse estudo.

REFERÊNCIAS

- ARAKAKI, F. A. **Linked Data**: ligação de dados bibliográficos. 2016. Nº f.144 Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências de Marília, Universidade Estadual Paulista, Marília, 2016. Disponível em: <https://repositorio.unesp.br/handle/11449/147979>. Acesso em: 06 abr. 2018.
- BERNERS-LEE, T. **Linked Data-design Issues**. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 01 jan. 2018.
- BIZER, C., HEATH, T., BERNERS-LEE, T. Linked Data: The Story so Far. *In: INTERNATIONAL JOURNAL ON SEMANTIC WEB AND INFORMATION SYSTEMS*. 5. 1-22. 2009. Disponível em: https://www.researchgate.net/publication/225070216_Linked_Data_The_Story_so_Far acesso em: 16/02/2018
- CHARLES, V., FREIRE, N., ANTOINE, I. Links, languages and semantics: linked data approaches in the European Library and Europeana. *In: SATELLITE MEETING ON LINKED DATA IN LIBRARIES*. Paris. **Anais...** Paris: IFLA, 2014. Tema: Linked Data in Libraries: Let's make it happen!
- CONEGLIAN, C. S.; SANTAREM, J. E. S. Europeana no Linked Open Data: conceitos de web semântica na dimensão aplicada das humanidades digitais. **Encontros Bibli**, Florianópolis, v. 22, n. 48, p. 88-99, 2017. Disponível em: <https://goo.gl/XTG56C>. Acesso em: 27 jul. 2018.
- EUROPEANA. **Europeana Data Model Primer**. 2011. Disponível em: <https://pro.europeana.eu/resources/standardization-tools/edm-documentation#primer>. Acesso em: 02 jun. 2017.
- HASLHOFER, B.; ISAAC, A. Data. europeana. eu: the europeana linked open data pilot. *In: I INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS*. 2011, **Anais...** 2011. p. 94-104. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3625>. Acesso em 21/04/2018.
- ISOTANI, S; BITTENCOURT, I. **Dados abertos conectados**: em busca da web do conhecimento. São Paulo: Novatec, 2015.
- LIRA, M. A. B. de. **Uma abordagem para enriquecimento semântico de metadados para publicação de dados abertos**. 2014, Nº95f, Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de Pernambuco, Escola de Ciência da Informação, Recife, 2014 Disponível em: <https://repositorio.ufpe.br/bitstream/handle/123456789/11570/DISSERTA%C3%87%C3%83O%20M%C3%A1rcio%20Angelo%20Bezerra%20de%20Lira.pdf?sequence=1&isAllowed=y> Acesso em: 22/12/2017
- NAZÁRIO, C. E. **Proposta de metodologia para avaliar o enriquecimento semântico de objetos publicados na web através do Linked Data**. 2018. 138f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2018.
- PITA, M.; PAIXAO, G. Arquitetura de busca semântica para governo eletrônico. *In: II WORKSHOP DE COMPUTAÇÃO APLICADA EM GOVERNO ELETRÔNICO & CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO*, 2., 2010, Belo Horizonte. **Anais...** Belo Horizonte: WCGE, 2010. p. 648-661.
- PITA, Marcelo; PAIXAO, Goedson. Arquitetura de Busca Semântica para Governo Eletrônico. *In, Belo Horizonte*. 2010.
- SILVA, D. L. da. **Ontologias para representação de documentos multimídia**: análise e modelagem. 2014. 442f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas

Gerais, Belo Horizonte, 2014. Disponível em:

<http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/BUOS-9NCGYM>. Acesso em: 24 out. 2016.

SILVA, W. D. da. Anotação semântica automática do Currículo Lattes utilizando Linked Open Data. **Projetos e Dissertações em Sistemas de Informação e Gestão do Conhecimento**, Belo Horizonte, v. 4, n. 1, 2016.

Disponível em: <http://www.fumec.br/revistas/sigc/article/view/2924/1754>. Acesso em: 24 nov. 2016.

SORRENTINO, S. *et al.* Semantic annotation and publication of Linked Open Data. *In: INTERNATIONAL CONFERENCE - ICCSA, 13., 2013, Ho Chi Minh City . Anais...* Ho Chi Minh City, Vietnam: [s.n.], 2013. p. 462-474.

TORINO, E. *et al.* Enriquecimento semântico para a disponibilização de dados abertos: teoria e prática.

Encontros Bibli. Florianópolis, v. 25, p. 1-19, 2020. Disponível em:

<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2020.e67991>. Acesso em: 13 abr. 2020.

TRIQUES, M. L.; ARAKAKI, A. C. S. Representação de patrimônios culturais em plataformas digitais: o modelo de dados da Europeia. **Em Questão**, Porto Alegre, v. 27, n. 1, p. 185-209, 2021.

UREN, V. *et al.* Semantic annotation for knowledge management: requirements and a survey of the state of the art. **Journal of Web Semantics**, p.14-28, 2006.