

# Proposta Metodológica para Avaliar o Enriquecimento Semântico de Objetos Publicados em *Linked Data*

*Methodological Proposal to Evaluate the Semantic Enrichment of Published Objects in Linked Data*

Claudiane Emanuele Nazário (1), Célia da Consolação Dias (2)

(1) (2) Universidade Federal de Minas Gerais - Escola de Ciência da Informação, [cnazario@gamarratecnologia.com.br](mailto:cnazario@gamarratecnologia.com.br); [celiadias@gmail.com](mailto:celiadias@gmail.com)

## Resumo

Linked Data é um conjunto de princípios proposto por Tim Berners Lee, cuja finalidade é a de facilitar a publicação e conexão de dados de diferentes fontes na Web Semântica. Atualmente vários modelos de dados são desenvolvidos com o objetivo de realizar o enriquecimento semântico desses dados para sua publicação em Linked Data, garantindo assim a interoperabilidade e a integração de informações de diferentes provedores. O presente artigo tem como objetivo apresentar uma proposta de metodologia para avaliar o enriquecimento semântico de objetos publicados na web através do Linked Data, utilizando uma Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos - Matriz TRESO desenvolvida durante a pesquisa de mestrado em Ciência da Informação. A Matriz desenvolvida foi aplicada nos modelos de dados BIBFRAME e EDM para verificar como esses modelos realizavam o enriquecimento semântico de objetos durante a publicação em Linked Data. Foi realizada uma análise comparativa através da qual foi possível identificar o modelo mais aderente aos critérios da Matriz TRESO, bem como propor recomendações para o enriquecimento semântico.

**Palavras-chaves:** Linked Data; Enriquecimento Semântico; Anotação Semântica.

## Abstract

Linked Data is a set of principles proposed by Tim Berners Lee, whose purpose is to facilitate the publication and connection of data from different sources in the Semantic Web. Currently, several data models are developed with the purpose of semantic enrichment of this data for publication in Linked Data, thus ensuring the interoperability and integration of information from different providers. The present article aims to present a methodology proposal to evaluate the semantic enrichment of objects published on the web through Linked Data, using a Matrix of Techniques and Resources for the Semantic Enrichment of Objects - Matrix TRESO developed during the master's research in Science of Information. The developed matrix was applied in the BIBFRAME and EDM data models to verify how these models performed the semantic enrichment of objects during the publication in Linked Data. A comparative analysis was carried out through which it was possible to identify the model most closely adhering to the criteria of the TRESO Matrix, as well as to propose recommendations for semantic enrichment.

**Keywords:** Linked Data; Semantic Enrichment; Data Models; Semantic Annotation.

## 1. Introdução

A Web Semântica surgiu com o objetivo de facilitar o compartilhamento de informações pelos usuários, atribuindo significado ao conteúdo existente. Alinhada às necessidades e às tecnologias demandadas pela *Web Semântica*, surgiu o conjunto de princípios propostos por Tim Bernes Lee denominado "*Linked Data*", cuja finalidade é a publicação e a conexão de dados estruturados de diferentes fontes na web, garantindo a interoperabilidade entre sistemas. (BERNERS-LEE, 2006).

A proposta do *Linked Data* consiste na conexão de dados através de links semânticos significativos também para o computador, automatizando tarefas antes possíveis somente a humanos. (BIZER et al., 2009)

Atualmente são publicados vários modelos, cuja proposta é a realização do enriquecimento semântico dos dados para sua publicação em *Linked Data*, de

modo a permitir a integração de dados de diferentes provedores.

Este artigo apresenta parte da pesquisa de mestrado do Programa de Pós-Graduação em Gestão e Organização do Conhecimento (PPGGOC) da Universidade Federal de Minas Gerais (UFMG), cujo objetivo geral foi desenvolver uma metodologia capaz de avaliar o enriquecimento semântico de objetos publicados na web através do *Linked Data*. Para tanto, foi desenvolvida uma Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos (Matriz TRESO) que foi aplicada em dois modelos de dados a saber: EDM (*Europeana Data Model*) e BIBFRAME (*Bibliographic Framework Initiative*) para testar a metodologia proposta.

O EDM foi desenvolvido pela Europeia para 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 e ISAAC, 2012, CHARLES, ISAAC e FREIRE, 2014)

E o BIBFRAME, desenvolvido pela *Libray of Congress* tem por objetivo implementar um novo ambiente para as bibliotecas na web, que favoreça a interconexão de informações e o intercâmbio de dados. (LIBRARY OF CONGRESS, 2012).

Como objetivos específicos da pesquisa citada foram propostos:

- Estudar os aspectos que favorecem o enriquecimento semântico de dados;
- Identificar como os modelos de dados fazem a publicação de dados em *Linked Data*;
- Realizar um estudo comparativo entre dois modelos de dados, utilizando uma Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos.

A seção a seguir apresenta parte do referencial teórico utilizado na pesquisa.

## 2. Referencial Teórico

O enriquecimento semântico é um processo utilizado para inserir conceitos a partes de um texto, de modo a criar uma estrutura para busca automática de seu significado. (THAKKER et al., 2012)

Lira (2014) afirma que o enriquecimento semântico é um processo de atribuição de maior significado aos dados e metadados, tornando os mesmos mais qualificados, através do uso da semântica atribuída por vocabulários pré-existentes, sinônimos e informações de proveniência, de modo a facilitar a compreensão, a integração e o processamento dos dados por homens e máquinas.

Uma das técnicas mais utilizadas no processo de enriquecimento semântico é a anotação. Segundo Lira (2014) a anotação semântica consiste na adição de informações de forma manual ou automatizada para atribuir significado aos recursos, isto é, anexar dados em partes de outros dados.

Bloehdorn, et al. (2005) apresentam os requisitos da anotação semântica, enfatizando a necessidade de utilização de ontologias de domínio no processo de anotação em objetos multimídia, de modo a melhorar a qualidade da anotação e permitir a criação de inferências.

Para Isotani e Bittencourt (2015) as relações semânticas criadas pelas ontologias podem ser utilizadas pelo *Linked Data* com vistas a ampliar as conexões entre os recursos na web. A partir do enriquecimento semântico às necessidades dos usuários são traduzidas por agentes de software, que avaliam a requisição original e, utilizando inferências lógicas, formadas a partir das ontologias e das conexões criadas pelo *Linked Data*, podem descobrir outros recursos na web de maneira semântica e não sintática. (ISOTANI; BITTENCOURT, 2015).

A seção a seguir apresenta a Matriz Técnicas e Recursos para o Enriquecimento Semântico de Objetos – Matriz TRESO para sua publicação em *Linked Data* desenvolvida para avaliar os modelos de dados EDM e BIBFRAME.

## 3. Construção da Matriz TRESO

Para atingir os objetivos do estudo identificou-se a necessidade de verificar quais os recursos e as técnicas são requeridos para a implementação do enriquecimento semântico. A Matriz TRESO disponível no Apêndice A foi criada para avaliar quais os recursos e técnicas estão presentes nos modelos de dados para o enriquecimento semântico. Esta matriz foi desenvolvida a partir de elementos identificados na literatura, considerando os estudos de LIRA (2014), SILVA (2016), SILVA (2014), PITA e PAIXÃO (2010), UREN et al. (2006) e SORRENTINO et al. (2013).

Na construção da Matriz TRESO foram considerados os seguintes critérios de avaliação:

1. Anotação semântica realizada. Neste critério é avaliado se o modelo de dados utiliza o recurso de anotação semântica e, em caso positivo identificar o tipo de anotação realizada. Neste caso, pode ser: anotação por *tags*; atributos/propriedades; anotação por meio de relações; anotação semântica utilizando elementos do vocabulário controlado e anotação semântica utilizando ontologias.
2. Reuso de metadados: Este critério avalia se o modelo de dados reutiliza dados e metadados na publicação de dados em *Linked Data*, otimizando o trabalho de publicação e se esta reutilização ocorre em alguma classe específica ou em todas as classes do modelo.
3. Links entre os dados e metadados do modelo com outros recursos da web: Este critério avalia se o modelo de dados faz combinações semânticas de seus dados e metadados com algum *dataset* específico ou com quaisquer *datasets* da web de dados, como o Geonames e a Dbpedia.
4. Modelagem de dados num formato semântico estruturado: Este critério avalia se a estruturação do modelo é realizada de forma padronizada, ampliando as possibilidades de conexão com outros *datasets* do *Linked Open Data* (LOD).
5. Utilização de ferramentas para o enriquecimento semântico: Este critério avalia se o modelo utiliza alguma ferramenta tecnológica para facilitar o processo de enriquecimento semântico.
6. Utilização de interface gráfica: Este critério avalia se o modelo utiliza alguma interface gráfica

para facilitar o processo de enriquecimento semântico e publicação de dados.

7. Relações entre os metadados do modelo e outros vocabulários: Este critério avalia se o modelo utiliza relações de sinonímia (equivalência), associação e hierarquia entre o metadado e o termo correspondente em outros vocabulários utilizados para facilitar o entendimento do usuário e reduzir a ambiguidade.

Os critérios foram organizados na Matriz TRESO e inicialmente foram estabelecidos os parâmetros para pontuação, considerando uma escala de 0 (zero) a 4 (quatro), em que cada modelo de dados seria pontuado de acordo com a forma de atendimento de cada critério, conforme descrito na Apêndice A–. Matriz TRESO

A seção a seguir apresenta o estudo comparativo entre os modelos.

#### 4. Estudo comparativo entre os modelos de dados EDM e BIBFRAME

Esse estudo consistiu em identificar se os modelos EDM e BIBFRAME atendiam aos recursos e técnicas para o enriquecimento semântico presentes na Matriz TRESO proposta.

Cada modelo foi avaliado de acordo com os critérios definidos na Matriz TRESO e pontuado de acordo com o grau de atendimento a cada quesito.

Após a pontuação dos modelos em cada critério o resultado foi somado para verificar a aderência dos mesmos à proposta de enriquecimento semântico.

O resultado obtido em cada modelo foi levantado para identificar os pontos comuns e os divergentes em cada um destes modelos, conforme os resultados apresentados na seção 4.1 a seguir.

##### 4.1 Resultados da Comparação dos Modelos de Dados

A Tabela 1 a seguir apresenta a pontuação obtida pelos modelos EDM e BIBFRAME em cada um dos critérios propostos na Matriz TRESO

Tabela 1 – Pontuação dos modelos

CRITÉRIO	EDM	BIBFRAME
Anotação semântica	3	2
Reuso de metadados	4	0
Links entre o modelo com outros recursos da web.	4	2
Modelagem de dados	4	3

CRITÉRIO	EDM	BIBFRAME
em formato semântico estruturado.		
Utilização de ferramentas para o enriquecimento semântico.	3	2
Utilização de interface gráfica.	0	0
Relacionamento entre os metadados do modelo e termos de outros vocabulários.	4	2
<b>TOTAL</b>	<b>22</b>	<b>9</b>

Fonte: elaboração própria

Os resultados dessa comparação e as justificativas para a pontuação obtida pelos modelos em cada um dos critérios são apresentados a seguir.

##### 4.1.1 Critério 1 - Anotação semântica

O primeiro critério da Matriz TRESO em que os modelos EDM e o BIBFRAME foram avaliados se refere ao modo com que cada modelo adiciona informações semânticas para atribuir significado aos recursos.

Após a análise observou-se que, ambos os modelos utilizam o recurso de anotação para enriquecimento semântico de objetos. O EDM permite a utilização de links e vocabulários, atribuindo maior qualidade da anotação, se comparado ao BIBFRAME que, realiza a anotação por meio de *tags* e de propriedades que permitem o uso de texto livre e linguagem natural.

Como o EDM não adota o recurso de anotação utilizando ontologias, considerado na pesquisa superior em relação aos demais tipos de anotação não obteve pontuação total.

##### 4.1.2 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.

Conforme apresentado na versão 2.0 do BIBFRAME, Library of Congress (2016), as propriedades *identifier*, *language*, *source* e *title*, possuem a mesma nomenclatura e função, presentes no padrão Dublin Core, formato de metadado básico endereçado a descrever recursos na Web. Deste modo, se trata de uma duplicação e não da reutilização da propriedade em seu vocabulário de origem.

A propriedade *CopyrightRegistration* possui a mesma finalidade da propriedade *Rights management* do Dublin Core, ou seja, o mesmo dado é referenciado de modo distinto em diferentes vocabulários, o que dificulta a reutilização e, sobretudo a integração dos dados no *Linked Data*. (LIBRARY OF CONGRESS, 2011).

No EDM 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. (EUROPEANA, 2011).

#### 4.1.3 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 dos modelos BIBFRAME e 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 BIBFRAME apresenta *links* apenas para os vocabulários do RDF (*Resource Description Framework*), modelo para representação de recursos na Web expresso a partir da tripla sujeito-predicado-objeto, Bizer et al., (2009) e do Dublin Core. Já o EDM realiza a conexão com vários outros vocabulários disponíveis na Web de Dados.

O quadro 1 apresenta um comparativo entre conexões do BIBFRAME e EDM segundo o LOV – *Linked Open Vocabularie*, portal da Web que apresenta os vocabulários publicados no LOD.

Quadro 1 – Conexões do EDM e BIBFRAME

CRITÉRIO	BIBFRAME	EDM
Utilização do Vocabulário no LOD	0 <i>datasets</i>	10 <i>datasets</i>
Links de saída	2 links	13 links
Links de entrada	1 link	1 link
Classes utilizadas	0 propriedades	11 classes
Propriedades utilizadas	0 propriedades	35 propriedades

Fonte: elaboração própria

Em relação a este quesito, foi possível observar que o EDM atende totalmente ao critério estabelecido em virtude dos seguintes pontos:

- a) O modelo realiza conexões com vários *datasets* do *Linked Open Data* (LOD). De

acordo com Arakaki (2016), o objetivo do EDM é o de justamente de utilizar o LOD para padronizar a descrição dos recursos do patrimônio cultural de instituições, como bibliotecas, arquivos, museus e galerias.

- b) O EDM prevê a integração entre as combinações semânticas dos dados e metadados, utilizando ontologias para conectar com quaisquer *datasets* publicados no LOD.

Já o BIBFRAME não atingiu nota máxima neste quesito por se conectar com apenas dois *datasets* do LOD.

#### 4.1.4 Critério 4 - Modelagem de dados em formato semântico estruturado

No que se refere ao critério Modelagem de dados num formato semântico estruturado, tanto o EDM, quanto o BIBFRAME são modelados por meio do RDF/XML, ou seja, representam o RDF por meio da linguagem *Extensible Markup Language* (XML). (EUROPEANA, 2011, LIBRARY OF CONGRESS, 2011).

No BIBFRAME as relações existentes entre os objetos e suas propriedades são descritos por meio de grafos RDF o que favorece uma maior liberdade e simplificação das representações. No entanto, neste modelo não são definidas a natureza e a tipologia das relações, ou seja, não é explicitado se as relações são do tipo todo-parte e gênero-espécie.

O EDM também utiliza a linguagem OWL (*Web Ontology Language*) linguagem para publicação de ontologias recomendada pela W3C para representar o significado dos termos e seus relacionamentos em vocabulários. (HECKMANN et. al., 2005).

O EDM também utiliza os conceitos da ontologia Cidoc Conceptual Reference Model (CIDOC CRM) no processo de modelagem. Segundo Carrasco, Thaller e Vidotti (2015, p. 211), o CIDOC CRM “é uma ontologia do domínio do patrimônio cultural, criada como uma ferramenta para integração da informação”. A utilização de ontologias resulta em uma maior qualidade dos dados e dos relacionamentos entre os mesmos. O uso da ontologia CIDOC CRM e suas generalizações pelo EDM têm por objetivo tanto modelar os dados, quanto realizar consultas em grandes volumes de dados.

#### 4.1.5 Critério 5 - Utilização de ferramentas para o enriquecimento semântico

Ambos os modelos utilizam ferramentas para o enriquecimento semântico de objetos. O BIBFRAME fornece uma ferramenta em Excel para a conversão de dados em formato MARC (*Machine-Readable Cataloging*). Segundo Alves e Souza (2007) o MARC é um conjunto de padrões para identificar, armazenar e

comunicar informações bibliográficas em formato legível por computador, permitindo que diferentes sistemas possam reconhecer, processar e estabelecer pontos de acesso dos elementos que compõem a descrição bibliográfica.

Segundo a Europeia (2014), a instituição desenvolveu uma pequena ferramenta tecnológica de enriquecimento para fazer a vinculação de dados, de modo totalmente automático. São realizadas referências entre os objetos do EDM com as URI (*Uniform Resource Identifier*), identificadores únicos de recursos na internet de outros vocabulários externos, bem como são recuperados dados multilíngues e semânticos relacionados a um determinado conceito do vocabulário.

Apesar de o EDM possuir ferramentas para apoiar no processo de enriquecimento semântico de forma automática, o fato das ferramentas serem proprietárias dificulta sua utilização pelos usuários, em virtude disto não foi atribuída a pontuação total, uma vez que as anotações ou vinculação de dados são realizadas por meio de ferramentas proprietárias e não por meio de software livre.

Para o BIBFRAME foi atribuída nota 2 – baixo atendimento, porque a ferramenta em formato Excel disponibilizada pelo modelo realiza somente a conversão do formato do MARC para o BIBFRAME, de forma semiautomática por publicadores e usuários.

#### 4.1.6 Critério 6 - Utilização de interface gráfica

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

Segundo Uren et al. (2005) a utilização de ferramentas automatizadas para o enriquecimento semântico é fundamental para reduzir os gaps na aquisição de conhecimento, principalmente considerando as grandes coleções de documentos existentes na web.

Neste critério, foi atribuída a nota zero – critério inexistente a ambos os modelos, tendo em vista que não foi possível identificar em nenhum dos mesmos a utilização de interface gráfica para o enriquecimento semântico.

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

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

No BIBFRAME as relações semânticas não são hierárquicas, associativas e de equivalência como algumas que ocorrem em uma ontologia, e sim representadas por meio de grafos RDF. Segundo a Library of Congress. (2012), o uso de grafos RDF permite

uma maior liberdade e simplificação das representações. No entanto, este requisito não atende ao critério proposto pela Matriz TRESO.

Já no EDM apesar das conexões entre os vocabulários serem realizados por meio do RDF, é possível identificar os tipos de relacionamentos existentes entre os objetos, através de propriedades relacionadas no quadro a seguir:

Quadro 2 - Tipologia dos Relacionamentos do EDM

Tipo de Relacionamento	Descrição	Propriedade correspondente
Hierarquia	Propriedades representadas por relações (todo-parte; gênero - espécie ou espécie - tipo)	dcterms:hasPart, dcterms:isPartOf, edm:hasType, edm:type
Associação	Propriedades relacionadas a outras (similar a; ver também)	edm:isSimilarTo
Sinonímia ou equivalência	Propriedades idênticas as outras	owl:sameAs

Fonte: Elaboração própria

Neste item foi atribuída a pontuação total ao EDM, 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.

Para o BIBFRAME foi atribuída **nota zero – critério inexistente** já que o mesmo é modelado em RDF e não explicita o tipo de relacionamento.

#### 4.2.1 Pontuação Geral dos Modelos.

Em relação ao resultado da avaliação, observou-se na Tabela 1- Pontuação dos modelos que o EDM, obteve 22 (vinte e dois) pontos, sendo classificado como atendimento total aos requisitos. Já o BIBFRAME obteve 9 (nove) pontos, sendo classificado como atendimento parcial, Quadro 1.

O resultado encontrado aponta que enquanto o BIBFRAME se encontra em desenvolvimento no que se refere à aplicação de técnicas e recursos para o enriquecimento semântico de objetos para sua publicação em *Linked Data*, o EDM se encontra em um nível mais avançado. Esse resultado pôde ser

comprovado na prática, na medida em que, enquanto o BIBFRAME não é utilizado por nenhum *dataset* do LOD, o EDM é utilizado em 10 (dez) desses *datasets*.

Deste modo, vale ressaltar alguns elementos fundamentais para amparar a enriquecimento semântico de objetos:

- a) A utilização de vocabulários controlados e ontologias para facilitar a compreensão da anotação pelo usuário.
- b) A reutilização de metadados de outros modelos em seu contexto original.
- c) A conexão com outros recursos da web.
- d) A modelagem de dados em um formato semântico estruturado, permitindo sua compreensão por homens e máquinas e, sua consequente reutilização.
- e) A explicitação dos relacionamentos entre os metadados do modelo e termos de outros modelos.
- f) E a utilização de ferramentas para facilitar o trabalho do publicador e permitir a manipulação de grandes volumes de dados.

Apesar de o modelo EDM possuir a maior parte dos elementos propostos pela Matriz TRESO para o enriquecimento semântico de objetos para sua publicação em *Linked Data*, entende-se que o mesmo ainda pode incorporar melhorias em seu modelo, tais como: uso de interface gráfica para agilizar o trabalho de publicação de grandes volumes de dados e utilizar ontologias no processo de enriquecimento semântico. Com relação ao BIBFRAME, entende-se que uma revisão das propriedades atualmente utilizadas no modelo, privilegiando sempre que possível à reutilização de metadados de outros modelos, bem como a utilização de ontologias e vocabulários para modelagem de seus dados pode contribuir para alinhar este modelo ao conceito de enriquecimento semântico. Outro item que poderia trazer benefícios ao BIBFRAME é a utilização de ferramentas tecnológicas para facilitar o processo de enriquecimento semântico.

## 5. Considerações Finais

O presente artigo buscou apresentar a pesquisa de mestrado cujo objetivo foi ampliar a discussão sobre o enriquecimento semântico de objetos para sua publicação em *linked data*.

Iniciou-se o estudo buscando suporte na literatura para compreender a cerca dos conceitos de modelagem de dados, do enriquecimento semântico, e dos princípios para publicação de dados em *Linked Data*.

Para realizar a pesquisa foi necessário desenvolver uma metodologia de análise que permitisse avaliar os

modelos de dados, uma vez que não foi identificado um instrumento de coleta de dados que possibilitasse verificar se de fato os modelos de dados continham todos os elementos para favorecer o enriquecimento semântico. Deste modo, durante a pesquisa foi desenvolvida a Matriz TRESO, inserida como apêndice neste artigo.

Acredita-se que os resultados obtidos foram satisfatórios e tornam viáveis à continuidade da pesquisa iniciada, na medida em que novos modelos de dados podem ser desenvolvidos ou aprimorados a partir da metodologia proposta.

Ressalta-se que as discussões acerca do comportamento dos modelos de dados, em cada um dos critérios elencados na Matriz TRESO, indicaram a possibilidade de, em trabalhos futuros, utilizar este instrumento 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*.

Deste modo, espera-se que a Matriz TRESO seja utilizada não só para avaliar outros modelos de dados, mas também para auxiliar na construção de novos modelos voltados para a publicação de dados em *Linked Data* enriquecidos semanticamente.

Ainda em relação às pesquisas futuras, pretende-se avançar no estudo dos modelos de dados e de recomendações para facilitar a publicação de dados em *Linked Data*. Pretende-se também aplicar a Matriz TRESO em modelos de outros domínios do conhecimento e possivelmente na construção de um *dataset* de patrimônio cultural desenvolvido a partir das recomendações da Matriz TRESO.

## Referências

ALVES, M; SOUZA, M. I. F. **Estudo de correspondência de elementos metadados:** Dublin Core e Marc 21. Embrapa Informática Agropecuária-Artigo em periódico indexado (ALICE), 2007. Disponível em: <https://www.alice.cnptia.embrapa.br/bitstream/doc/1328/1/APEstudocorrespondenciaAlveseSouza2007pdf.pdf>

ARAKAKI, F. A. et al. **Linked Data:** ligação de dados bibliográficos. 144 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016. Disponível em: <https://alsafi.ead.unesp.br/handle/11449/147979>. Acesso em: 13 jul. 2017.

BERNERS-LEE, T. **Linked Data-design issues.** 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 15 mar. 2017.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data-the story so far. **Semantic services, interoperability and web applications: emerging concepts**, United States of America, p. 205-227, 2009. Disponível em: <[https://books.google.com.br/books?hl=ptBR&lr=&id=tP8HLEtgbKcC&oi=fnd&pg=PA205&dq=Linked+Data:+Design+issues&ots=-hJuToD1BB&sig=dT6uRAuVjm\\_XBZ\\_iCWa-9EWLmP4#v=onepage&q=Linked%20Data%3A%20Design%20issues&f=false](https://books.google.com.br/books?hl=ptBR&lr=&id=tP8HLEtgbKcC&oi=fnd&pg=PA205&dq=Linked+Data:+Design+issues&ots=-hJuToD1BB&sig=dT6uRAuVjm_XBZ_iCWa-9EWLmP4#v=onepage&q=Linked%20Data%3A%20Design%20issues&f=false)>. Acesso em: 15 mar. 2017.

BLOEHDORN, S. et al. **Semantic annotation of images and videos for multimedia analysis**. In: EUROPEAN SEMANTIC WEB CONFERENCE, 2nd, 2005, Heraklion. Proceedings... Heraklion: [s.n.], 2005.

CARRASCO, L. B.; TALLHER, M.; VIDOTTI, S. A. B. G. **Ontologia Cidoc CRM no contexto dos ambientes digitais de patrimônios culturais**. Liinc em Revista, Rio de Janeiro, v.11, n.1, p. 208-222, maio 2015. Disponível em: <<http://liinc.revista.ibict.br/index.php/liinc/article/view/789>>. Acesso em: 19 nov. 2016.

CHARLES, V.; ISAAC, A.. **Europeana and (many) linked open vocabularies**. In: LOV Symposium: Linking and Opening Vocabularies, Universidad Carlos III de Madrid, Spain, 2012. Slides.

CHARLES, V; ISAAC, A; FREIRE, N. **Linking Libraries in The European Library e Europeana**. In: IFLA 2014 SATELLITE MEETING LINKED DATA IN LIBRARIES, 2014, [S.l.]. **Anais...** [S.l.], IFLA, 2014.

EUROPEANA. **EDM definition: 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](http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM%20Definition%20v5.2.6_01032015.pdf)>. Acesso em: 2 mar. 2017.

EUROPEANA. **Europeana Data Model Primer**. 2011. Disponível em: <<https://pro.europeana.eu/resources/standardization-tools/edm-documentation#primer>>. Acesso em: 02 jun. 2017

HECKMANN, D.; SCHWARTZ, T.; BRANDHERM, B.; SCHMITZ, M.; VON WILAMOWITZ-MOELLENDORFF, M. **Gumo: The General User Model Ontology**. In: USER MODELING. 10. 2005, Berlin. Proceedings... Heidelberg: Springer-Verlag, 2005. p. 428-432.

ISOTANI, S.; BITTENCOURT, I. I. **Dados Abertos Conectados: Em busca da web do Conhecimento**. Novatec Editora, 2015.

LIBRARY OF CONGRESS. **Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services**. Washington, DC, 2012. Disponível em: <<http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>>

LIBRARY OF CONGRESS. **BIBFRAME vocabulary: terminology and conventions**. EUA: LC, 2011. Disponível Em: <<http://www.loc.gov/bibframe/docs/vocab-conventions.html>>.

LIBRARY OF CONGRESS. **Overview of the BIBFRAME 2.0 model**. Washington, DC, 2016. Disponível em <<https://www.loc.gov/bibframe/docs/bibframe2-model.html>>.

LIRA, M. A. B. de. **Uma Abordagem Para Enriquecimento Semântico de Metadados Para Publicação de Dados Abertos**. 2014. 95 f. Dissertação (Mestrado em Ciência da Informação) - Centro de Informática, Universidade Federal de Pernambuco, 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: 19 fev. 2018.

PITA, M.; PAIXAO, G. **Arquitetura de Busca Semântica para Governo Eletrônico**. In: WCGE, 2., 2010, Belo Horizonte. **Anais...** Belo Horizonte: WCGE, 2010.

SILVA, D. L. da. **Ontologias para representação de documentos multimídia: análise e modelagem**. 2014. 442 f. 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/sig/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.

THAKKER, D.; DESPOTAKIS, D.; DIMITROVA, V.; LAU, L.; BRNA, P. **Taming Digital Traces for Informal Learning: A Semantic-driven Approach**. In: EUROPEAN CONFERENCE ON TECHNOLOGY ENHANCED

LEARNING, 7., 2012, Berlin. Proceedings... Heidelberg: Springer-Verlag, 2012.

UREN, V., CIMIANO, P., IRIA, J., HANDSCHUH, S., VARGAS-VERA, M., MOTTA, E., CIRAVEGNA, F., **Semantic Annotation for Knowledge Management: Requirements and a survey of the state of the art.** Journal of Web Semantics: Science, Services and Agents on the World Wide Web 4, 14-28, 2006. Disponível em: <[https://nemo.inf.ufes.br/wp-content/papercite-data/pdf/documentacao semantica no apoio a integracao d e dados e rastreabilidade 2010.pdf](https://nemo.inf.ufes.br/wp-content/papercite-data/pdf/documentacao_semantica_no_apoio_a_integracao_d_e_dados_e_rastreabilidade_2010.pdf)>. Acesso em: 12 out. 2016.

## Apêndice A - Matriz TRESO

(Continua)

Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos		Modelo de dados: ( ) EDM ( ) BIBFRAME Data: ____/____/____			
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 anotações 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> .
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 dataset específico.	Existência de links entre as combinações semânticas dos dados e metadados com até 2 datasets específico.	O modelo prevê a integração entre as combinações semânticas dos dados e metadados com quaisquer datasets da web de dados, como o Geonames e a Dbpedia.	O modelo prevê a integração entre as combinações semânticas dos dados e metadados, utilizando ontologias para conectar com quaisquer datasets 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 de dados, ampliando as possibilidades de conexão com outros Datasets 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

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 de outros proprietários 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: Elaboração própria