

UNIVERSIDADE FEDERAL DE MINAS GERAIS

ESCOLA DE CIÊNCIA DA INFORMAÇÃO

CLAUDIANE EMANUELE NAZÁRIO

**PROPOSTA DE METODOLOGIA PARA AVALIAR O  
ENRIQUECIMENTO SEMÂNTICO DE OBJETOS PUBLICADOS NA  
WEB ATRAVÉS DO *LINKED DATA***

Belo Horizonte

2018

CLAUDIANE EMANUELE NAZÁRIO

**PROPOSTA DE METODOLOGIA PARA AVALIAR O  
ENRIQUECIMENTO SEMÂNTICO DE OBJETOS PUBLICADOS NA  
WEB ATRAVÉS DO *LINKED DATA***

Dissertação apresentada ao programa de pós-graduação em Gestão & Organização do Conhecimento (PPG-GOC) da Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Ciência da Informação

Linha de Pesquisa: Arquitetura & Organização do Conhecimento.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup> Célia da Consolação Dias

Belo Horizonte

2018

N335p

Nazário, Claudiane Emanuele.

Proposta de metodologia para avaliar o enriquecimento semântico de objetos publicados na web através do Linked Data [manuscrito] / Claudiane Emanuele Nazário. – 2018.

140 f., enc. : il.

Orientadora: Célia da Consolação Dias.

Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f. 121-127.

Apêndices: f. 128-140.

1. Ciência da informação – Teses. 2. Web Semântica – Teses. 3. Representação do conhecimento (Teoria da informação) – Teses. 4. Modelagem de dados – Teses. I. Título. II. Dias, Célia da Consolação. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 004.82



## FOLHA DE APROVAÇÃO

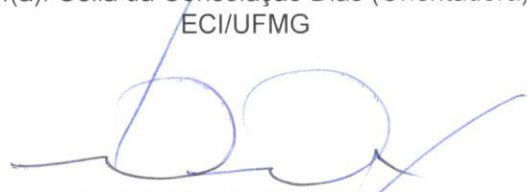
PROPOSTA DE METODOLOGIA PARA AVALIAR O ENRIQUECIMENTO SEMÂNTICO DE  
OBJETOS PUBLICADOS NA WEB ATRAVÉS DO LINKED DATA

### CLAUDIANE EMANUELE NAZARIO

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Arquitetura e Organização do Conhecimento.

Aprovada em 09 de maio de 2018, pela banca constituída pelos membros:

  
Prof(a). Célia da Consolação Dias (Orientadora)  
ECI/UFMG

  
Prof(a). Renato Rocha Souza  
FGV/RJ (por videoconferência)

  
Prof(a). Gercina Ângela de Lima  
ECI/UFMG

Belo Horizonte, 9 de maio de 2018.



## ATA DA DEFESA DA DISSERTAÇÃO DA ALUNA CLAUDIANE EMANUELE NAZARIO

Realizou-se, no dia 09 de maio de 2018, às 10:00 horas, Sala 1000 - ECI/UFMG, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *PROPOSTA DE METODOLOGIA PARA AVALIAR O ENRIQUECIMENTO SEMÂNTICO DE OBJETOS PUBLICADOS NA WEB ATRAVÉS DO LINKED DATA*, apresentada por CLAUDIANE EMANUELE NAZARIO, número de registro 2016662535, graduada no curso de BIBLIOTECONOMIA/DIURNO, como requisito parcial para a obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Celia da Consolação Dias - ECI/UFMG (Orientadora), Prof(a). Renato Rocha Souza - FGV/RJ (por videoconferência), Prof(a). Gercina Ângela de Lima - ECI/UFMG.

A Comissão considerou a dissertação:

Aprovada

Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 09 de maio de 2018.

  
Prof(a). Celia da Consolação Dias

  
Prof(a). Renato Rocha Souza

  
Prof(a). Gercina Ângela de Lima

## **AGRADECIMENTOS**

Agradeço a Deus por ter me suportado durante esta caminhada e me ungido com sua imensa sabedoria.

A meu pai Jonas que apesar de todas às dificuldades me fez entender a importância da educação para superar todas as barreiras que a vida me apresentou.

A minha mãe Alecilda que me apoiou e incentivou nas horas difíceis, de desânimo e cansaço.

Aos meus filhos Rafael e Luiz Henrique que entenderam a minha ausência e sempre se orgulharam das minhas conquistas.

A meu marido Roberto e minha irmã Flaviane com quem sempre posso contar incondicionalmente.

Aos meus irmãos Robson, Jackson, Thiago, Gisele, Jonathan e Bryan que sempre foram grandes incentivadores.

As minhas grandes amigas, Gerilza, Renata, Carla e Dayara que sempre me estimularam.

As minhas colegas de trabalho Mamede, Maria Fernanda, Sandra, Karina e Elisângela pela experiência e inspiração.

Agradeço aos professores Maurício, Bax, Renato, Gercina e todos os demais, não somente por terem me ensinado, mas por terem me feito aprender.

E finalmente, agradeço a minha orientadora Celia da Consolação Dias que com toda sua paciência, dedicação, competência e bom humor fez surgir em mim o desejo de me tornar pesquisadora.

## RESUMO

A Web Semântica surgiu para facilitar o compartilhamento de informações pelos usuários, atribuindo significado ao conteúdo existente. Para auxiliar neste processo surgiu o “*Linked Data*”, um conjunto de princípios cuja finalidade é facilitar a interoperabilidade entre os sistemas. Contudo, os diferentes vocabulários utilizados pelas diversas fontes de informação e a ausência de padrões para publicação dos dados acaba por dificultar o processo de representação do conhecimento e qualidade dos dados disponibilizados na web. Esta pesquisa tem por objetivo o enriquecimento semântico de objetos publicados na web através do *Linked Data*. Como metodologia de trabalho foi realizado um levantamento bibliográfico para a construção do referencial teórico metodológico deste estudo. Na Revisão da Literatura foram apresentados conceitos, ferramentas e tecnologias relativas ao tema. A construção dos fundamentos teóricos metodológicos permitiu a criação de uma **Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos** - Matriz TRESO. A matriz criada foi aplicada nos modelos de dados BIBFRAME e EDM para verificar se os modelos de dados aplicavam o enriquecimento semântico de objetos durante a publicação em *Linked Data*. Como resultados deste estudo foi proposta uma metodologia para avaliar o enriquecimento semântico de objetos publicados em *Linked Data*.

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

## ABSTRACT

The Semantic Web has emerged to facilitate the sharing of information by users, assigning meaning to existing content. To assist in this process came the "*Linked Data*", a set of principles whose purpose is to facilitate interoperability between systems. However, the different vocabularies used by the various sources of information and the absence of standards for data publication end up hampering the process of representing the knowledge and quality of the data made available on the web. This research aims at the semantic enrichment of objects published on the web through *Linked Data*. As a work methodology, a bibliographic survey was carried out to construct the theoretical methodological reference of this study. In the Review of Literature were presented concepts, tools and technologies related to the theme. The construction of the theoretical methodological foundations allowed the creation of a Matrix of Techniques and Resources for the Semantic Enrichment of Objects - Matrix TRESO. The created matrix was applied in the BIBFRAME and EDM data models to verify if the data models applied semantic enrichment of objects during the publication in *Linked Data*. As results of this study, a methodology was proposed to evaluate the semantic enrichment of objects published in *Linked Data*.

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



## LISTA DE FIGURAS

|   |     |
|---|-----|
| Figura 1 - The Semantic Web Tecnology Stack .....   | 39  |
| Figura 2 - The Linking Open Data cloud diagram..... | 45  |
| Figura 3 - Etapas do Ciclo de Vida do LOD2.....     | 47  |
| Figura 4 - Estrutura da URI.....                    | 49  |
| Figura 5 - Estrutura URL, URI, URN .....            | 50  |
| Figura 6 - Exemplo de grafo RDF .....               | 52  |
| Figura 7 - Tela de consulta SPARQL .....            | 57  |
| Figura 8 - Resultado da consulta SPARQL .....       | 58  |
| Figura 9 – Metodologia .....                        | 66  |
| Figura 10 - Utilização do BIBFRAME no LOV .....     | 106 |
| Figura 11 - Utilização do EDM no LOV .....          | 107 |

## LISTA DE GRÁFICOS

|  |     |
|--|-----|
| Gráfico 1 - Evolução dos Datasets.....   | 46  |
| Gráfico 2 – Anotação semântica entre EDM e BIBFRAME .....                        | 100 |
| Gráfico 3 - Reutilização de Metadados EDM e BIBFRAME .....                       | 103 |
| Gráfico 4 – Link entre dados e metadados do EDM e BIBFRAME com recursos da Web.. | 108 |
| Gráfico 5 – Modelagem de dados em um formato estruturado EDM e BIBFRAME .....    | 110 |
| Gráfico 6 – Utilização de ferramentas para o Enriquecimento semântico.....       | 112 |
| Gráfico 7 – Relações entre o metadado e os termos de outros vocabulários .....   | 114 |

## LISTA DE QUADROS

|  |    |
|--|----|
| Quadro 1 - Tipos de Anotações Semânticas .....   | 29 |
| Quadro 2 – Requisitos das ferramentas de anotação semântica .....  | 33 |
| Quadro 3 - Transcrição da Consulta SPARQL .....  | 56 |
| O quadro 4 a seguir apresenta as sete classes do EDM:Quadro 4 - Classes do EDM.....                      | 63 |
| Quadro 5 - Classes do BIBFRAME .....   | 65 |
| Quadro 6 - Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos –<br>versão 1.0..... | 73 |
| Quadro 7 - Formulário de Avaliação de Enriquecimento semântico .....                                     | 77 |
| Quadro 8 - Validação da Matriz no BIBFRAME.....  | 80 |
| Quadro 9 - Validação da Matriz no EDM.....   | 81 |
| Quadro 10 - Escala de Pontuação por critério.....  | 82 |
| Quadro 11 - Escala de Pontuação Total da Matriz .....  | 83 |
| Quadro 12 - Matriz TRESO versão 2.0 .....  | 85 |
| Quadro 13 – Aplicação da Matriz ao BIBFRAME .....  | 91 |
| Quadro 14 – Aplicação da Matriz no EDM .....   | 94 |
| Quadro 21 – Aplicação da Matriz no EDM .....   | 94 |

|   |     |
|---|-----|
| Quadro 15 - Índice de Reutilização das propriedades do EDM .....  | 102 |
| Quadro 17 – Comparativo entre as conexões do EDM e BIBFRAME no LOV .....  | 105 |
| Quadro 18 - Tipologia dos Relacionamentos do EDM.....   | 113 |
| Quadro 19 - Propriedades do Objeto Patrimônio Cultural (Provided Cultural Heritage Object (edm:ProvidedCHO) ..... | 126 |
| Quadro 20 - Propriedades da classe Recursos da web - Web Resource (edm:webResource) .....                         | 129 |
| Quadro 21 - Propriedades para a classe aggregation (ore:Aggregation).....   | 130 |
| Quadro 22 - Propriedades para a classe Agent (edm:Agent) .....  | 132 |
| Quadro 23 - Propriedades para a classe Place (edm:Place).....   | 133 |
| Quadro 24 - Propriedades para a classe Timespan ou Period (edm:TimeSpan).....                                     | 133 |
| Quadro 25 - Propriedades para a classe Concept (skos:Concept).....  | 135 |
| Quadro 26 - Propriedades do BIBFRAME.....   | 135 |

## LISTA DE SIGLAS E ABREVIATURAS

|          |  |
|----------|--|
| BCI      | Biblioteconomia e Ciência da Informação                      |
| BIBFRAME | Bibliographic Framework Initiative                           |
| EDM      | Europeana Data Model   |
| FRBR     | Functional Requirements for Bibliographic Records            |
| HTTP     | Hypertext Transfer Protocol                                  |
| KOS      | Knowledge Organization Systems                               |
| LOD      | Linked Open Data   |
| LOV      | Linked Open Vocabularie                                      |
| OWL      | Web Ontology Language  |
| RDF      | Resource Description Framework                               |
| TRESO    | Técnicas e Recursos para Enriquecimento Semântico de Objetos |
| SGBD     | Sistema de Gerenciamento de Banco de Dados                   |
| SOC      | Sistemas de Organização do Conhecimento                      |
| SPARQL   | Structure Protocol and RDF Query Language                    |
| URI      | Uniform Resource Identifier                                  |

## SUMÁRIO

|           |   |           |
|-----------|---|-----------|
| <b>1</b>  | <b>INTRODUÇÃO .....</b>                                       | <b>16</b> |
| 1.1       | Problema de Pesquisa e Justificativa .....                    | 17        |
| 1.2       | Objetivo Geral .....  | 19        |
| 1.3       | Objetivos Específicos .....                                   | 19        |
| 1.4       | Estrutura da Dissertação .....                                | 19        |
| <b>2</b>  | <b>FUNDAMENTOS TEÓRICOS METODOLÓGICOS.....</b>                | <b>21</b> |
| 2.1       | Representação do Conhecimento .....                           | 21        |
| 2.1.1     | Modelagem de dados.....                                       | 23        |
| 2.1.2     | Modelagem conceitual .....                                    | 25        |
| 2.1.3     | Enriquecimento Semântico.....                                 | 26        |
| 2.1.3.1   | Anotação Semântica.....                                       | 28        |
| 2.1.3.1.1 | Tipos de anotação semântica .....                             | 28        |
| 2.1.3.1.2 | Requisitos das Ferramentas de Anotação Semântica .....        | 32        |
| 2.1.3.2   | Uso de ontologias para Busca Semântica .....                  | 34        |
| 2.1.3.3   | Vinculação de Dados e Modelos Orientados a Objetos .....      | 36        |
| 2.2       | <i>Linked Data</i> .....                                      | 36        |
| 2.2.1     | Padrões e modelos para publicação em <i>Linked Data</i> ..... | 48        |

|          |   |           |
|----------|---|-----------|
| 2.2.1.1  | URI .....   | 48        |
| 2.2.1.2  | Modelo RDF .....  | 51        |
| 2.2.1.3  | OWL .....   | 54        |
| 2.2.1.4  | SPARQL.....   | 55        |
| 2.3      | Modelos de Dados .....  | 59        |
| 2.3.1    | Europeana Data Model (EDM) .....  | 59        |
| 2.3.2    | Bibliographic Framework Initiative (BIBFRAME).....  | 63        |
| <b>3</b> | <b>METODOLOGIA.....</b>   | <b>66</b> |
| 3.1      | Caracterização da Pesquisa.....   | 66        |
| 3.2      | Pesquisa bibliográfica .....  | 67        |
| 3.3      | Subsídios teóricos para a construção da matriz proposta .....                                       | 67        |
| 3.4      | Delimitação da Pesquisa .....   | 69        |
| 3.5      | Procedimentos Metodológicos .....   | 69        |
| <b>4</b> | <b>CONSTRUÇÃO DA MATRIZ DE TÉCNICAS E RECURSOS PARA O ENRIQUECIMENTO SEMÂNTICO DE OBJETOS .....</b> | <b>70</b> |
| 4.1      | 4.1 CRITÉRIO DE AVALIAÇÃO .....   | 71        |
| <b>5</b> | <b>RESULTADOS E ANÁLISE .....</b>   | <b>79</b> |
| 5.1      | Validação da matriz.....  | 79        |
| 5.1.1    | Validação da matriz no Modelo BIBFRAME .....  | 79        |

|       |  |            |
|-------|--|------------|
| 5.1.2 | Validação da matriz no Modelo EDM .....  | 80         |
| 5.2   | Estudo comparativo entre os modelos de dados EDM e BIBFRAME .....                              | 90         |
| 5.2.1 | Aplicação dos Critérios no BIBFRAME .....  | 90         |
| 5.2.2 | Aplicação dos Critérios da Matriz ao EDM .....   | 94         |
| 5.3   | Resultados da Comparação dos Modelos de Dados .....  | 98         |
| 5.3.1 | Critério 1 - Anotação semântica .....  | 99         |
| 5.3.2 | Critério 2 - Reuso de metadados.....   | 100        |
| 5.3.3 | Critério 3 - Links entre dados e metadados do modelo com outros recursos da web                | 104        |
| 5.3.4 | Critério 4 - Modelagem de dados num formato semântico estruturado.....                         | 108        |
| 5.3.5 | Critério 5 - Utilização de ferramentas para o enriquecimento semântico .....                   | 110        |
| 5.3.6 | Critério 6 - Utilização de interface gráfica .....   | 112        |
| 5.3.7 | Critério 7 – Relacionamento entre os metadados do modelo e termos de outros vocabulários. .... | 113        |
|       | <b>6 CONSIDERAÇÕES FINAIS.....</b>   | <b>117</b> |
|       | <b>REFERÊNCIAS .....</b>   | <b>119</b> |
|       | <b>APÊNDICE .....</b>  | <b>126</b> |



# 1 INTRODUÇÃO

A Web Semântica tem como objetivo facilitar o compartilhamento de informações pelos usuários, atribuindo significado ao conteúdo existente. Um dos desafios para a sua implementação, no entanto está em identificar as relações existentes entre os vários recursos disponibilizados na internet, para que possam ser codificados, manipulados e disseminados por computador.

A Web Semântica também pode ser entendida como resultado das iniciativas para agregar valor aos dados disponibilizados na web, de modo a torná-los mais acessíveis aos homens e máquinas. Trata-se de uma extensão da web tradicional, onde as páginas são enriquecidas por uma estrutura semântica, cuja compreensão do conteúdo presente nestas páginas possa ocorrer tanto por agentes humanos, quanto por agentes de software. (ISOTANI e BITTENCOURT, 2015)

Segundo Halpin e Laverenko (2009) a Web Semântica é o resultado da aplicação de tecnologias de representação de conhecimento em sistemas distribuídos, com o objetivo de eliminar lacunas de comunicação existentes entre o homem e a máquina.

Diferente da web tradicional onde as páginas são interligadas por meio de hiperlinks, a Web Semântica emprega tecnologias como o RDF (*Resource Description Framework*) para publicar dados estruturados na web. É o uso de tais tecnologias que torna possível conectar diferentes fontes, de modo a permitir que uma imensa quantidade de dados existentes na internet possa ser acessada e compreendida não somente por pessoas, mas também por máquinas (BATISTA; LOSCIO, 2013).

Alinhado aos conceitos, às necessidades e às tecnologias demandadas pela *Web Semântica*, surgiu o conjunto de práticas propostas por Tim Bernes Lee denominado de "*Linked Data*", cuja finalidade é publicar dados estruturados na web, interligando dados de diferentes fontes, permitindo e facilitando assim a interoperabilidade entre sistemas. A proposta do *Linked Data* consiste em interligar dados abertos por meio de *links* semânticos significativos também para programas de computador, de modo a automatizar tarefas e

conexões antes possíveis somente por humanos.

Atualmente são publicados muitos modelos de dados, cuja proposta é justamente realizar o enriquecimento semântico desses dados para sua publicação em *Linked Data*, garantindo a interoperabilidade e integração de dados de diferentes provedores.

A presente pesquisa buscou propor uma metodologia para avaliar o enriquecimento semântico de objetos publicados na web através do *Linked Data*. Para tanto, foram avaliados dois modelos de dados para validar a metodologia proposta neste estudo.

## 1.1 Problema de Pesquisa e Justificativa

Atualmente há uma quantidade grande de repositórios e bases de dados sobre diferentes domínios de conhecimento (medicina, direito, educação etc.) e a disponibilização de conteúdo em vários formatos (música, mídia, publicações) disseminados diariamente na web. No entanto, na maioria das vezes, os dados são disseminados na internet sem adotar qualquer padrão ou controle terminológico e tal situação pode dificultar a sua interpretação por sistemas informatizados.

A falta de um vocabulário controlado, elemento próprio dos tesouros, das taxonomias e das ontologias, acabam por dificultar a integração e a interoperabilidade de dados advindos de fontes heterogêneas, bem como a publicação de dados *Linked Data*.

Marcondes e Campos (2008) afirmam que as primeiras tentativas de utilizar computadores no processamento semântico de conteúdo datam das décadas de 1950 e 1960, por pesquisadores da Biblioteconomia e Ciência da Informação (BCI) como Shera (1957) e Luhn (1960). Já faziam parte das discussões de teóricos desde esta época, as possibilidades dos computadores criarem inferências ou axiomas - a partir da análise de outros relacionamentos presentes na modelagem conceitual, e a tentativa de ampliar o processamento automático de conteúdo.

Apesar dos avanços do uso de *Linked Data* em outras áreas do conhecimento

como a computação e a engenharia, observa-se que a Ciência da Informação ainda carece de estudos sobre a temática e que essa área poderia contribuir no desenvolvimento do assunto.

As discussões no âmbito dos estudos sobre o *Linked Data*, não abrangem de forma substancial as contribuições da BCI para o desenvolvimento de modelos, metodologias e ferramentas voltadas à representação, tratamento, busca e recuperação da informação, tais como: modelos de dados, vocabulários controlados e padrões para descrição e recuperação de informação.

O uso do “*Linked Data*” associado às técnicas de enriquecimento semântico contribuem tanto para a otimização da qualidade de dados disponibilizados, quanto para facilitar a representação do conhecimento e a interoperabilidade entre diferentes fontes de dados.

A literatura apresenta diferentes modelos de dados que realizam o enriquecimento semântico de objetos, visando sua publicação no *Linked Data*, contudo não foi identificada uma forma para avaliar se os recursos e técnicas utilizados por estes modelos para realizar o enriquecimento semântico realmente contribuem para a qualificação destes dados.

Deste modo, a questão investigada nesse estudo foi como avaliar o enriquecimento semântico de objetos publicados na web através do *Linked Data*? A resposta a esse questionamento teve como foco contribuir tanto para a implementação de melhorias, no que se refere à semântica e qualidade dos dados publicados na web, quanto para possibilitar a compreensão da importância do processo de modelagem conceitual, com vistas a melhorar a qualidade das informações recuperadas em domínios diversos e favorecer a interoperabilidade dos dados.

## 1.2 Objetivo Geral

Desenvolver uma metodologia para avaliar o enriquecimento semântico de objetos publicados na web através do *Linked Data*.

## 1.3 Objetivos Específicos

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

## 1.4 Estrutura da Dissertação

Esta dissertação foi organizada em seis capítulos. O primeiro capítulo apresenta o contexto de realização deste estudo, possibilitando uma visão geral da temática, bem como situando a questão de pesquisa, a justificativa e os objetivos.

O segundo capítulo descreve a fundamentação teórico-metodológica que consiste em três seções: A primeira seção apresenta uma visão sobre a representação do conhecimento, a modelagem de dados, a modelagem conceitual e o enriquecimento semântico. A segunda e a terceira seção desse capítulo apresentam, respectivamente os conceitos atinentes ao *Linked Data* e a apresentação dos modelos de dados utilizados nesta pesquisa.

O terceiro capítulo aborda o desenvolvimento da dissertação com o detalhamento da caracterização da pesquisa, revisão de literatura, delimitação da pesquisa e os procedimentos metodológicos.

O quarto capítulo apresenta a proposta de metodologia para avaliação do enriquecimento semântico de objetos para publicação através do *Linked Data*,

compreendendo a criação da **Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos – Matriz TRESO**.

O quinto capítulo apresenta a validação da Matriz TRESO desenvolvida na etapa anterior e a aplicação desse instrumento nos dois modelos de dados analisados nesta pesquisa, os resultados alcançados com a implementação da matriz.

O sexto capítulo traz as considerações finais para o estudo realizado e as possibilidades para trabalhos futuros.

## 2 FUNDAMENTOS TEÓRICOS METODOLÓGICOS

Neste capítulo apresenta-se um panorama teórico-contextual acerca dos conceitos e temas que fundamentaram essa pesquisa. Expõem-se as ideias de diferentes autores que corroboraram no início da dissertação com os insumos necessários para uma discussão qualitativa sobre o estudo que foi realizado.

Inicia-se este capítulo com as contribuições do campo de estudo da representação do conhecimento, sobretudo no que se refere à modelagem conceitual, à modelagem de dados e ao enriquecimento semântico. A seguir são apresentados conceitos, padrões, modelos e tecnologias para a publicação de dados em *Linked Data*. Finalizando o capítulo apresentam-se os modelos de dados EDM e BIBFRAME que foram utilizados para testar os objetivos dessa pesquisa.

### 2.1 Representação do Conhecimento

O conhecimento foi fixado e inserido através do processo de comunicação e dos elementos da linguagem. De acordo com Dahlberg, (1978) a linguagem pode ser definida como a capacidade humana de denominar os objetos do mundo que o rodeia, bem como a habilidade do homem de se comunicar com os seus semelhantes. É a linguagem que torna possível formular enunciados a respeito dos objetos do mundo, de modo a torná-los compreensíveis.

A linguagem contribuiu sobremaneira para o processo de representação do conhecimento, na medida em que através dela foi possível para os seres humanos pensar nas características, atribuir rótulos e conceitos aos objetos existentes no mundo.

Segundo Alvarenga (2003) representar refere-se ao “ato de colocar algo no lugar de”, e, também, que a representação do conhecimento é a expressão simbólica da observação da natureza e dos fatos sociais, utilizando as linguagens disponíveis no contexto da comunicação e da produção do conhecimento.

A representação [...] destaca-se como uma instância do processo cognitivo humano, aquela que culmina com a representação primária do conhecimento, situando-se no âmbito do registro do pensamento em um suporte documental, incluindo as etapas de percepção, identificação, interpretação, reflexão e codificação, etapas que são envolvidas no ato de se conhecer um novo ser ou coisa, ou aprofundar-se no conhecimento de um ser ou uma coisa já conhecida, utilizando-se dos sentidos, da emoção, da razão e da linguagem. [...] (ALVARENGA, 2003, p.21).

Os instrumentos de representação da informação utilizados para organizar o conhecimento de determinado domínio de forma semântica são denominados sistemas de organização do conhecimento (SOC), ou do inglês Knowledge Organization Systems – (KOS). Souza; Tudhope; Almeida (2012). São exemplos de SOC os tesauros, taxonomias e ontologias, na medida em que são utilizados para organizar e representar o conhecimento através de sistemas informatizados.

O termo Knowledge Organization Systems (KOS) foi utilizado inicialmente por Hodge (2000) para classificar todos os de esquemas de organização da informação e gerenciamento do conhecimento, tais como: listas de cabeçalhos de assunto, tesauros, dicionários, glossários, ontologias dentre outros.

Segundo Soergel, 2014, os KOS cobrem uma grande variedade de sistemas que atendem a uma ampla gama de finalidades. Estes sistemas são conhecidos como ontologia, esquema de metadados, taxonomia, classificação, estrutura de diretórios da Web, plano de arquivo, dicionário de sinônimos, folksonomia dentre outros.

Ainda, sobre estes sistemas Aguiar e Kobashi (2013) descrevem que os SOC são formas de materializar o processo de Representação do Conhecimento, servindo de meio de comunicação entre os produtores e usuários da informação. Para Tristão et. al. (2004) os diversos SOC tinham como objetivo a organizar, recuperar e disseminar a informação para os usuários que dela necessitassem.

No que se refere à organização e recuperação de documentos digitais, além das etapas de registro, classificação, indexação e armazenamento também existe a necessidade de estabelecer e extrair metadados diretamente do conteúdo destes documentos. (ALVARENGA, 2003).

Ainda, segundo Alvarenga (2003), as facilidades disponibilizadas pelo meio digital o tornam um espaço sem precedentes para o registro e a recuperação de documentos textuais, sonoros e imagéticos devido à enorme gama de possibilidades de armazenagem, memória e formatos. No entanto, este espaço acaba por demandar novos elementos para promover a recuperação desses registros de conhecimento na web.

Se por um lado o uso intensivo de tecnologias facilita a publicação de um número cada vez maior de documentos digitais, por outro, a representação do conhecimento desses documentos demanda a utilização de novos recursos e técnicas para permitir seu processamento de modo automático e dinâmico.

Para Marcondes (2011) a representação do conhecimento deve situar-se entre dois extremos, sendo: a) economizar energia de usuários e processadores da informação: b) ser suficientemente rica sob o aspecto cognitivo e compreensível para aqueles que dela necessitam.

Também, segundo Marcondes (2011) os modelos de representação do conhecimento, com seus conceitos e aplicabilidade, juntamente com seus benefícios surgem com a função de permitir que o usuário, ao final da leitura ou da visualização, possa inferir na verdadeira pertinência da informação diante de suas necessidades de modo mais econômico e preciso.

A seção 2.1.1 apresenta os conceitos de modelagem conceitual e seu papel na representação do conhecimento.

Hyland et al., (2014) consideram a etapa de modelagem como sendo uma das boas práticas para a publicação de dados abertos conectados, ou seja, a forma como os dados serão representados e como eles se relacionam com outros dados, independente da aplicação.

### **2.1.1 Modelagem de dados**

O contexto informacional pode ser caracterizado por uma grande quantidade de dados, e de termos que representam os conceitos existentes nos objetos de um domínio. Os



dados deste contexto precisam ser organizados e da mesma forma os termos que fazem parte de um vocabulário necessitam ser tratados e controlados.

A modelagem de dados consiste em explorar estruturas orientadas a dados. Através da modelagem de dados é possível identificar tipos de entidades, e atributos associados a estas entidades.

Sayão e Marcondes (2008) defendem que para a recuperação, reutilização e integração de diferentes dados na internet, são necessárias tecnologias que permitam alto grau de interoperabilidade. Contudo, a não utilização de modelos para a representação da informação, a ambiguidade presente na linguagem natural, bem como a falta de consenso quanto à terminologia utilizada em certo domínio, e muito frequentemente a ausência de um vocabulário que controle todos estes aspectos aqui mencionados, acabam por não favorecer a integração dos dados por mais robustos que sejam os sistemas utilizados.

Existem diferentes abordagens para a modelagem de dados. Essas abordagens podem ser divididas em quatro categorias:

- a) Relacionais: quando o modelo objeto da modelagem é baseado no conceito matemático de relação que utiliza uma coleção de tabelas para a representação dos conjuntos de dados e suas relações;
- b) Entidade-Relacionamento: quando o modelo utiliza objetos para explicitar, respectivamente: as entidades (objetos do mundo real) e seus relacionamentos (com outras entidades) através do uso de chaves;
- c) Orientados a Objetos: utilizado na construção de aplicativos por linguagens de programação como a C e o Java este modelo é baseado no aporte de Orientação a Objetos;
- d) Semiestruturado: modelo com características contidas em conjuntos de dados de mesma tipologia, com o uso de linguagens de marcação *XML* ou *JavaScript Object Notation*.

Silva (2011) destaca que a modelagem de dados no ambiente digital busca representar e organizar o conhecimento, por meio da criação de estruturas de dados

eletrônicos como banco de dados com vistas a um conjunto de informações. Para iniciar a modelagem de dados, deve-se priorizar a etapa de modelagem conceitual e para a compreensão do significado. A seção a seguir apresenta as definições de modelagem conceitual.

### **2.1.2 Modelagem conceitual**

A modelagem conceitual para Gomes (2014) pode ser entendida como parte essencial no desenho de um sistema de informação, na medida em que é a modelagem conceitual que identifica a especificação das estruturas dos dados e seus relacionamentos.

Para Furgeri (2006) e Silva (2011) a modelagem conceitual ajuda a reduzir conflitos semânticos e melhorar a documentação de sistemas, tendo em vista que é uma importante ferramenta para a representação e para o auxílio do desenvolvimento dos conceitos.

A partir da modelagem conceitual é possível identificar e extrair dados de determinado domínio de conhecimento, compreender as relações entre os conceitos existentes e “verificar as possibilidades de ligação/separação semânticas entre os conceitos de um dado domínio” (Campos, 2004, p.23), permitindo sua interpretação por homens e máquinas. Nesse processo de formação da estruturação conceitual de um domínio, o uso de termos ambíguos que possam causar conflitos semânticos, não ocorre devido à identificação das relações entre os objetos presentes neste contexto específico.

A modelagem conceitual possibilita a redução de ambiguidades e, portanto, facilitando a representação e compreensão dos dados por homens e máquinas, de modo a facilitar o enriquecimento semântico de objetos publicados na Web em diferentes áreas do conhecimento como engenharia, medicina, agricultura, dentre outros.

A seção a seguir apresenta os principais conceitos e técnicas para o enriquecimento semântico.

### 2.1.3 Enriquecimento Semântico

A Semântica é o estudo do significado na linguagem Hurford e Heasley (1997). Já o enriquecimento semântico pode ser entendido como um processo para inserir conceitos a partes de um texto, de modo a criar uma estrutura para busca automática de seu significado. (THAKKER et. al., 2012)

A literatura apresenta a visão de autores como Moreira (2014); Lira (2014); Heckmann et. al., (2005); Pita e Paixão (2010); sobre o enriquecimento semântico e que será apresentada a seguir.

De acordo com Moreira (2014) através de técnicas de processamento de linguagem natural, ontologias e algoritmos de seleção e marcação de conteúdo é possível acrescentar significado e enriquecer semanticamente os registros coletados, de acordo com os domínios de conhecimentos registrados em um repositório de dados.

Para Lira (2014) 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. Ainda, segundo Lira (2014), o objetivo do enriquecimento semântico é facilitar a compreensão, a integração e o processamento dos dados por pessoas e máquinas. O autor também descreve alguns recursos e técnicas para realizar o Enriquecimento Semântico, como será apresentado:

- a) Anotação Semântica – Consiste na adição de informação semântica de forma manual ou automatizada para atribuir significado aos recursos, isto é, anexar dados em partes de outros dados.
- b) Vinculação e Mapeamento de Recursos: consiste na descoberta de links entre as combinações semânticas dos dados e metadados com outros recursos na web de dados.
- c) Conversão para modelos de dados semânticos: Consiste em modelar os dados, utilizando um formato semântico e estruturado, como o RDF/XML, permitindo sua manipulação por aplicações que consomem esses modelos de dados.

As ontologias possuem um importante papel nesse processo de enriquecimento semântico, na medida em que possibilitam modelagem de conceitos e relacionamentos de acordo com os perfis de usuários, possibilitando realizar inferências sobre estes conceitos (HECKMANN et. al., 2005).

Lira (2014, p.44) pontua algumas características que um conjunto de metadados enriquecidos deve contemplar, tais como os atributos semânticos, os *datasets*, os vocabulários padrões e demais descritas a seguir:

- (i) “Maior quantidade de atributos semânticos, que contenham um significado claro, capaz de descrever melhor os dados;
- (ii) Facilidade de interpretação e processamento do conteúdo dos datasets , uma vez que o metadado estará em linguagem estruturada, ou seja, definida de forma que pode ser processada por máquina;
- (iii) Termos de vocabulários padrões associados, que permitirão ao metadado se integrar a outros dados e recursos na web.”

(LIRA 2014, p.44)

Ressalta-se que neste estudo a definição de Lira (2014) para enriquecimento semântico norteará as análises que serão realizadas nessa pesquisa.

As características de metadados apresentadas por Lira (2014) e pontuadas neste trabalho, serão usadas como referência para avaliar o enriquecimento semântico dos objetos, adotado em cada um dos modelos de dados que serão analisados na etapa empírica deste estudo.

Além do estudo realizado por Lira (2014) já mencionado anteriormente, também serão utilizados como referência para a construção da **Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos** que foi implementada nessa pesquisa, os trabalhos de Moura (2009); Uren et. Al (2005); Bontcheva e Cunningham (2011); Pita e Paixão (2010); Hohenstein e Plessner (1996); Silva (2014); Silva (2015); Bloehdorn, Stephan et al (2005); Sorrentino et al. (2013) e Chebotko et al. (2005) e que serão detalhados nas seções a seguir.

### 2.1.3.1 Anotação Semântica

A anotação semântica é uma das técnicas mais utilizadas no processo de enriquecimento semântico. Bloehdorn, Stephan et al 2005 apresentam requisitos da Anotação Semântica enfatizando a necessidade de utilização de ontologias de domínio no processo de anotação semântica em objetos multimídia.

Silva (2016) apresenta um levantamento bibliográfico sobre Anotação Semântica, abrangendo conceitos, componentes, ferramentas, tecnologias e suas aplicações, com o objetivo de auxiliar na implementação do processo de enriquecimento semântico.

Silva (2014) apresenta vários modelos de anotação semântica, incluindo *tags*, atributos, relações e ontologias. A seção 2.3.1.1 a seguir, apresenta os tipos de anotação semântica propostos pelo autor.

#### 2.1.3.1.1 *Tipos de anotação semântica*

Silva (2014) descreveu quatro tipos de anotação semântica, sendo: anotação utilizando *tags*; anotação por meio de atributos; anotação por meio de relações e, anotação por meio de ontologias. O quadro 1 a seguir apresenta as definições, limitações e os benefícios dos diferentes tipos de anotação, sob o prisma do formalismo e riqueza estrutural, bem como da exigência requerida do usuário envolvido no processo de anotação.

### Quadro 1 - Tipos de Anotações Semânticas

(Continua)

| Modelo                          | Descrição   | Vantagem  | Desvantagem   |
|---------------------------------|---|---|---|
| Anotação utilizando <i>tags</i> | A anotação por <i>tag</i> é realizada inserindo uma palavra-chave não hierárquica ou termo-livre associado a um recurso, (pessoas, lugares, tópicos, datas etc).  | Qualquer usuário pode facilmente anotar um recurso web e encontrar por meio de consulta ou busca outros recursos que foram anotados utilizando a mesma <i>tag</i>   | A utilização de linguagem natural pode promover equívocos, ambiguidade e dificuldade de interpretação por homens e máquinas.  |
| Anotação por meio de atributos  | Consiste na anotação semântica de um recurso utilizando as propriedades, atributos deste recurso, como: autor, idioma, edição, ano de publicação, dentre outros, permitindo a identificação do recurso. | Este modelo é bem familiar para os usuários, uma vez que é utilizado tanto na web como em desktops. Esse modelo apresenta superioridade semântica em relação a anotação realizada por meio de <i>tags</i> , na medida em que o próprio atributo já explicita o relacionamento entre a anotação e o recurso anotado. | Além de requerer maior conhecimento do usuário para realizar este tipo de anotação, é necessário considerar que a semântica deste modelo ainda é limitada e não permite muitos relacionamentos entre os recursos. |

(continuação)

| Modelo                        | Descrição   | Vantagem   | Desvantagem   |
|-------------------------------|---|--|---|
| Anotação por meio de relações | Trata-se de uma extensão do modelo de anotação por atributos, permitindo o usuário interligar vários recursos por meio das ligações entre eles. | Este modelo apresenta como vantagem o fato de fornecer um modo de interligar vários recursos através de links (relações), permitindo ao usuário navegar de um recurso para outro através destes links. | A desvantagem deste modelo é que não considera relações de alto nível entre os recursos conectados, não demonstrando com clareza a natureza destas relações, que pode frustrar o usuário, face ao desconhecimento sobre os recursos e os tipos de relacionamentos existentes entre os mesmos. |

(conclusão)

| Modelo                          | Descrição   | Vantagem   | Desvantagem  |
|---------------------------------|---|--|--|
| Anotação por meio de ontologias | Nesse modelo as ontologias são utilizadas como suporte a anotações, permitindo ao usuário descrever e interligar recursos existentes por meio de qualificadores como conceitos, instâncias, propriedades e restrições mantidas entre tais recursos. | São vantagens deste modelo: Permitir as relações explícitas e significativas entre dados estruturados e não estruturados;<br>Eliminar a ambiguidade;<br>Permitir a interpretação dos relacionamentos, utilizando agentes de software;<br>Realizar o processamento automático de anotações. | As desvantagens deste modelo são notórias dada a dificuldade de manutenção pelos anotadores e mediante a dificuldades de Interação do usuário com o software e os dados. |

Fonte: Adaptado pela autora de Silva (2014)

Bontcheva e Cunningham (2011), ainda apresentam os desafios a serem enfrentados pelas ferramentas de anotação semântica manual, sendo:

- 1º. Apoiar referências a conceitos de ontologia através de URIs;
- 2º. Ser colaborativas e também baseadas na web para permitir que equipes distribuídas de anotadores compartilhem o trabalho;
- 3º. Ir além da anotação de classes e instâncias, e também suportar a anotação de valores de propriedades e relacionamentos;
- 4º. Suportar a anotação em relação a ontologias múltiplas e dimensionar para suporte das ontologias com muitas classes e relações.



- 5º. Suportar múltiplos formatos de documentos, indo além do HTML para PDF, XML, imagens e vídeo.

#### *2.1.3.1.2 Requisitos das Ferramentas de Anotação Semântica*

Bontcheva e Cunningham (2011), discutem acerca das ferramentas de anotação semântica manual no contexto dos principais requisitos, como suporte para diversos formatos de documentos, ontologias múltiplas e anotações colaborativas baseadas na Web.

Uren et al., (2005) formularam sete requisitos para serem considerados pelas ferramentas de anotação semântica, partindo de quatro perspectivas: as ontologias; os documentos; as anotações que vinculam as ontologias aos documentos e os usuários dos sistemas. Cada perspectiva sugere um ou mais requisitos, conforme apresentado no quadro a seguir:

## Quadro 2 – Requisitos das ferramentas de anotação semântica

(Continua)

| Requisito  | Descrição   |
|--|---|
| <b>Requisito 1 – Formatos padronizados</b>                                     | <p>Os sistemas de anotação semântica devem utilizar sempre que possível, formatos padronizados. Na medida em que estes padrões podem fornecer um mecanismo para integração entre recursos heterogêneos, propiciam o acesso simultâneo por usuários e organizações permitindo o compartilhamento de anotações. Dois padrões são requeridos sendo o OWL para a descrição de ontologias e o RDF para realizar as anotações.</p> <p>A utilização de formatos-padrão é pré-requisito para o compartilhamento de anotações.</p>   |
| <b>Requisito 2- usuário centrado / projeto colaborativo</b>                    | <p>A anotação pode tornar-se um gargalo se for feito exclusivamente por trabalhadores do conhecimento com muitas demandas. É crucial fornecer a esses, interfaces que simplifiquem o processo de anotação inseridas no contexto do trabalho cotidiano. Uma abordagem interessante, seria a disponibilização de um ambiente em que os usuários anotem documentos integrados a esse, com vistas a criação, leitura, compartilhamento e edição de informações. O design do sistema também precisa facilitar a colaboração entre os usuários, possibilitando que especialistas de diferentes campos possam contribuir e reutilizar documentos inteligentes.</p> |
| <b>Requisito 3 – Suporte a ontologia (ontologias múltiplas e atualizações)</b> | <p>Além de suportar formatos de ontologia apropriados, as ferramentas de anotação precisam ser capazes de suportar múltiplas ontologias. Ou as ontologias devem ser conectadas ou, ainda, as anotações devem declarar explicitamente quais ontologia a que se referem. Além disso, os sistemas terão de lidar com as mudanças feitas nas ontologias ao longo do tempo, como a incorporação de novas classes ou alterações nas classes existentes. Nesse caso, o problema é assegurar consistência entre ontologias e anotações referentes às mudanças de ontologia.</p>   |
| <b>Requisito 4 - suporte de documentos de formatos heterogêneos</b>            | <p>Os padrões da Web semântica para anotação tendem a assumir que os documentos a serem anotados estão em formatos nativos da Web, tais como o HTML e XML.</p> <p>Os documentos estarão em muitos formatos diferentes, incluindo arquivos de processadores, planilhas, arquivos gráficos e misturas complexas de diferentes formatos. Deste modo, lidar com documentos de múltiplos formatos é um pré-requisito para integrar a anotação com as práticas de trabalho existentes.</p>  |

(conclusão)

| Requisito   | Descrição  |
|---|--|
| <b>Requisito 5 - atualização de documentos (documento e consistência de anotação)</b> | As ontologias mudam às vezes, mas alguns documentos mudam muitas vezes. Um exemplo são os documentos de especificação do W3C, que passam por várias revisões. Este requisito diz respeito a manter a consistência das anotações quando um documento é revisado.<br><br>Os ambientes de anotação precisam ajudar os trabalhadores do conhecimento a manterem anotações apropriadas à medida que os documentos mudam.                  |
| <b>Requisito 6 - armazenamento de anotações</b>                                       | As anotações podem ser armazenadas, tanto separadamente do documento original, quanto como parte integrante do documento, que pode ser visto ou não de acordo com as preferências do leitor.   |
| <b>Requisito 7- automação</b>   | Outro aspecto para reduzir o gargalo na aquisição de conhecimento é a marcação automática de documentos para facilitar a anotação de grandes coleções de documentos de modo econômico. Para conseguir isso, a integração de tecnologias de extração de conhecimento para a anotação é vital. Essas tecnologias podem identificar automaticamente entidades que são instâncias de uma determinada classe e relações entre as classes. |

Fonte: Adaptado pela autora de Uren et al. (2005)

### 2.1.3.2 Uso de ontologias para Busca Semântica

De acordo com Pita e Paixão (2010) um sistema de busca semântica considera tanto a utilização de padrões estruturais sintáticos, quanto de padrões semânticos para ampliar as possibilidades de busca. Desta forma, a busca semântica se torna possível quando agentes de software são capazes de compartilhar ontologias, constituindo assim uma infraestrutura natural para operacionalização da Web Semântica.

Isotani e Bittencourt (2015) apresentam a importância dessa infraestrutura para a publicação de dados. Para esses autores, no *Linked Data* essa infraestrutura pode ser empregada na medida em que 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 as necessidades dos usuários são traduzidas por agentes de software, que se encarregarão de avaliar a requisição original e, utilizando inferências lógicas, formadas a partir das ontologias e das conexões criadas pelo *Linked Data*, poderão descobrir outros recursos na web de maneira semântica e não sintática, como acontece na web tradicional. (ISOTANI; BITTENCOURT, 2015)

Em seguida, abordam sistemas orientados a ontologia, semiautomáticos e automáticos, que consideram tipicamente como seu formato de saída, mas não as usam como um recurso de conhecimento durante a análise semântica. Então, uma série de abordagens avançadas de anotação semântica baseadas em ontologias são apresentadas e comparadas entre si. Uma ênfase particular é a escalabilidade - que tem a capacidade de processar milhões de documentos, e a personalização – que trata da facilidade de adaptação desses sistemas a novos domínios e / ou ontologias.

Pita e Paixão (2010) descrevem o uso de agentes para compartilhamento de ontologias, em que as necessidades dos usuários são repassadas para os agentes de software que se encarregarão de enriquecer semanticamente a requisição original, utilizando inferência lógica e outras técnicas de inteligência artificial para descobrir outros recursos da web para atender a requisição. Pita e Paixão explicitam como o agente de busca realiza o enriquecimento semântico, de modo a expandir semanticamente o resultado da consulta, com base na ontologia armazenada.

Moura (2009) destaca o trabalho realizado pelo grupo de pesquisa ONTOI (grupo responsável por várias publicações e orientações com foco no tema de interoperabilidade de ontologias), cujo foco de um dos trabalhos é a Gerência Semântica de Documentos. Neste trabalho foi realizada a anotação semântica de documentos antes de seu armazenamento em um SGBD (Sistema de Gerenciamento de Banco de Dados), baseado em uma ontologia de domínio, com o objetivo de melhorar o desempenho das consultas. O trabalho consiste em realizar o tratamento semântico de documentos antes de sua carga no SGBD, mapeando-os para os termos de uma ontologia de domínio, antes do momento da consulta pelos usuários.

Chebotko et al., (2005) afirmam que o desenvolvimento da Web Semântica depende da disponibilidade de ontologias e poderosas ferramentas de anotação. Os autores

apresentam uma ferramenta de anotação multimídia linguística baseada em ontologia, OntoELAN, que apresenta: (1) suporte para ontologias OWL; (2) o gerenciamento de perfis de linguagem, que permitem ao usuário escolher um subconjunto de termos ontológicos para anotação; (3) o gerenciamento de níveis ontológicos, que podem ser anotados com termos de perfil de linguagem e, termos ontológicos correspondentes; e (4) armazenamento de documentos de anotação OntoELAN em formato XML com base em ontologias de domínio e multimídia.

### 2.1.3.3 Vinculação de Dados e Modelos Orientados a Objetos

Sorrentino et al., (2013) apresentam o conceito de vinculação e mapeamento de recursos para interligação de dados e metadados com recursos na nuvem do Linked Open Data (LOD).

Hohenstein e Plesser (1996) apresentam o modelo de dados orientado a objetos proposto pelo ODMG (Object Database Management Group) usado para homogeneizar aplicações locais, através de esquemas relacionais. É proposto o uso de um modelo de dados semanticamente rico para restaurar relacionamentos e hierarquias de subtipos em um processo de enriquecimento semântico, tornando explícita qualquer semântica dada nas relações usando conceitos orientados a objetos.

A seção a seguir apresentará os conceitos, as tecnologias relacionadas ao conjunto de princípios para a publicação de dados conectados na Web Semântica, denominados por Tim Bernes Lee como *Linked Data*.

## 2.2 *Linked Data*

Em 2016, de acordo com um *white paper* publicado pela Cisco<sup>1</sup>, o tráfego global de dados na rede ultrapassou a marca de 1 *Zettabyte*, o que equivalente a 1 bilhão de

---

<sup>1</sup> Cisco - Tráfego global de dados na rede. Para mais detalhes acessar: [https://www.cisco.com/c/pt\\_pt/about/press/news-archive-2017/20170208.html](https://www.cisco.com/c/pt_pt/about/press/news-archive-2017/20170208.html)

terabytes. Esta quantidade de informações torna cada vez mais difícil a busca e a recuperação de dados de forma eficiente e precisa pelos seres humanos.

Além dos problemas relacionados, há que se considerar também, a quantidade de informações existente na internet, em que grande parte dos dados não está disponível de forma estruturada para facilitar sua compreensão, acesso e manipulação. Esta situação demanda uma maior utilização de agentes de software para gerenciar, coletar, modelar, padronizar e consumir esses dados de forma automática.

Há alguns anos, diversas empresas, governos e institutos de pesquisa têm empreendido esforços para disponibilizar dados e produzir tecnologias web voltadas para a produção e consumo de dados. O objetivo desses esforços é facilitar a descoberta de novos conhecimentos e agregar valor a qualquer informação disponibilizada na Internet (ISOTANI e BITTENCOURT, 2015).

As necessidades advindas da evolução da web, como manipular um grande volume de dados e facilitar o acesso à informação, fizeram com que Tim Bernes-Lee em 2006, recomendasse um conjunto de princípios e boas práticas para publicação de dados na web. Estes princípios denominados *Linked Data* têm 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).

Os princípios propostos por Tim Bernes-Lee podem ser assim resumidos:

1. Utilizar URIs (Uniform Resource Identifier) para denominar os recursos da web;
2. Utilizar URIs HTTP (Hypertext Transfer Protocol) para que os usuários possam buscar esses recursos;
3. Fornecer informação útil, utilizando os padrões RDF (Resource Description Framework), SPARQL (serviços de consulta que utilizam o protocolo HTTP), quando alguém procurar uma URI;
4. Incluir links para outras URIs, possibilitando a descoberta de novos recursos. (BERNERS-LEE, 2006, p.1).

Assim, como já foi apontado anteriormente em relação à *Web Semântica* essas práticas, também são fundamentadas em tecnologias e padrões internacionais

recomendados pela W3C<sup>2</sup> tais como: HTTP (*Hypertext Transfer Protocol*), URI (*Uniform Resource Identifier*) e RDF (*Resource Description Framework*) para permitir a leitura e interpretação dos dados conectados, de forma automática, utilizando agentes de software para fazer uso efetivo dos dados publicados na web. (ISOTANI e BITTENCOURT, 2015).

Segundo Berners-Lee (2006), o modo com que os documentos estavam estruturados na web tradicional, isto é, por meio de nós e links, fazia com que, apenas seres humanos pudessem entender o significado contido neles. Os links utilizados para relacionar os documentos não apresentavam nenhum tipo de característica que os diferenciasse, sendo impossível para as máquinas distinguírem o significado existente entre uma relação e outra, bem como acessar e obter significado dos documentos.

Enquanto na web tradicional o relacionamento entre os documentos é realizado por meio de links de hipertexto entre as páginas HTML, na Web Semântica o que vincula os dados são as ligações em RDF, na qual uma parte do dado é conectada a uma parte de outro dado. (SANTOS NETO et al., 2013).

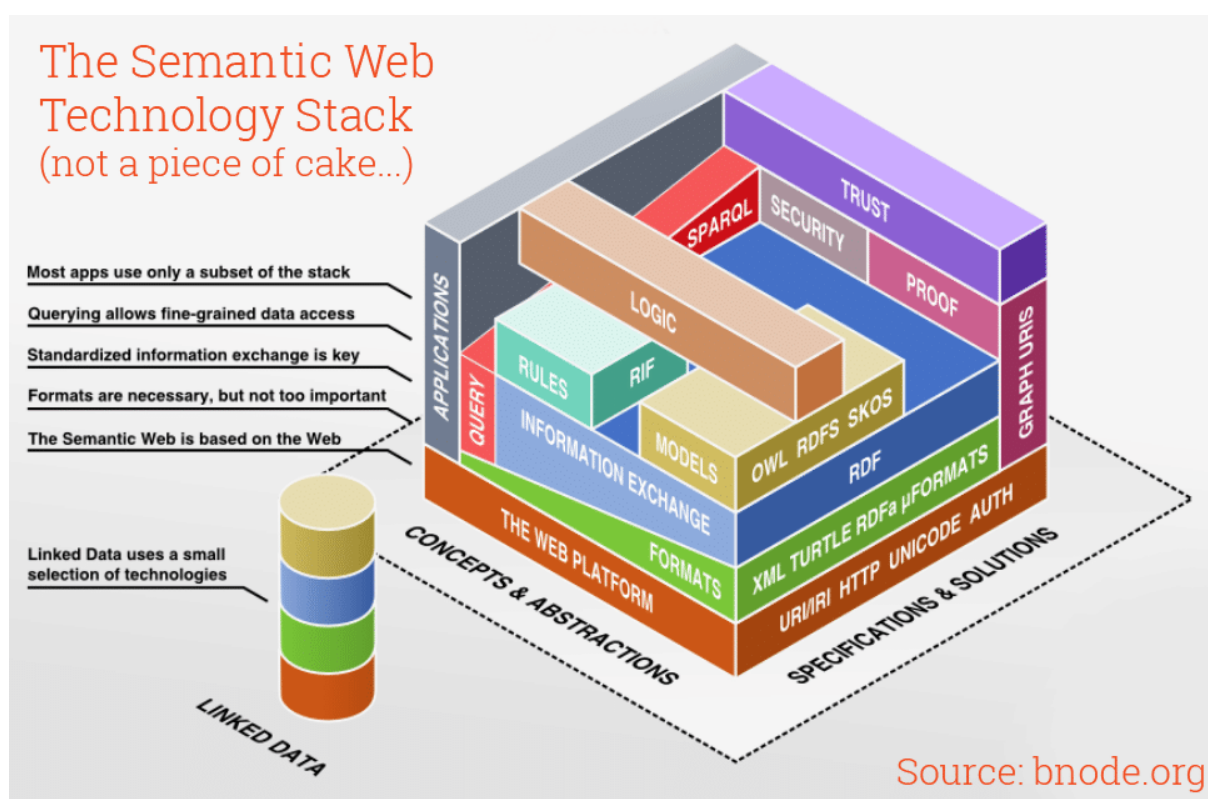
A utilização do *Linked Data*, com o uso do RDF, permite que a relação entre os recursos publicados na web se torne acessível também por máquinas. Desta forma, através de linguagens e padrões específicos, são criados links semânticos entre os recursos, possibilitando a execução automática de atividades, a exemplo do agendamento de consultas médicas, a compra de produtos, as reservas de hotéis, dentre outros serviços de modo personalizado.

A figura 01 **The Semantic Web Technology Stack** a seguir apresenta o conjunto de tecnologias e padrões adotados pela Web Semântica.

---

<sup>2</sup> W3C (*World Wide web Consortium*). Criado na Conferência Internacional de Genebra, em maio de 1994 com o objetivo de cuidar dos padrões e das tecnologias relacionados ao desenvolvimento da web.

Figura 1 - The Semantic Web Technology Stack



Fonte: BNODE (2018)

Berners-Lee (2006) aponta um conjunto de princípios para o *Linked Data* que considera: a) utilização de URI para identificar e denominar os recursos da web; b) adoção de URI HTTP para que os usuários possam localizar esses recursos; c) fornecimento de informação útil, utilizando os padrões RDF e SPARQL quando alguém procurar uma URI; d) inclusão de links para outras URIs, possibilitando a descoberta de novos recursos.

Segundo Pinheiro (2011), o sucesso da iniciativa do *Linked Data* pode ser justificado principalmente, em virtude de o mesmo apoiar-se em um conjunto de padrões amplamente utilizados na web, sendo:

- Um mecanismo de identificação global e único – URI;
- Um mecanismo de acesso universal HTTP;
- Um modelo de dados RDF,



d) Uma linguagem de consulta SPARQL para acesso aos dados.

Bizer e Heath (2011) acrescentaram aos princípios apresentados por Bernes-Lee alguns requisitos para publicação de dados em *Linked Data*, sendo eles:

- a) Os dados precisam estar no formato RDF;
- b) Devem ser desenvolvidas no mínimo 1.000 triplas;
- c) Deve ser conectado por no mínimo 50 links RDF para um conjunto de dados que já está no diagrama,
- d) Ser acessado por *SPARQL endpoint*, dentre outros.

O *Linked Data*, segundo Santos Neto, et al., (2013), surge para promover a integração de dados entre si e com outros dados, de modo a criar uma representação semântica legível por homens e máquinas. No entanto, ainda existem dificuldades para publicação de dados, adotando o *Linked Data*, em virtude de vários registros não serem disseminados na internet de modo aberto, padronizado e acessível.

A proposta do *Linked Data* está justamente em utilizar a web para propiciar ligações entre diversas fontes de dados. Como exemplos podem ser citados bancos de dados mantidos por organizações em localizações geográficas distintas, ou sistemas heterogêneos dentro de uma mesma organização que, não conseguem integrar seus dados de forma nativa, isto é, sem a necessidade de customizações. (BIZER; HEATH; BERNERS-LEE, 2009).

Para Isotani e Bittencourt (2015) existem dois desafios para a publicação dados em *Linked Data*. São eles: 1) escolher o vocabulário que melhor identifica o conjunto de dados e, 2) representar esses dados, de modo a aumentar sua expressividade em seu contexto de geração e reduzir a ambiguidade para sua posterior interpretação.

Observou-se que embora Heath e Bizer (2011) tenham mencionado sobre os requisitos para publicação de dados em *Linked Data*, eles não apresentaram explicações adicionais, especificamente sobre as quantidades mínimas para as triplas ou links necessários para a publicação de dados.

Da mesma forma Rocha (2012) afirma que o *Linked Data* utiliza a Web Semântica como plataforma de representação de informação, por meio de bases de dados compostas por triplas representadas na linguagem RDF, e o significado destas triplas especificado através de ontologias expressas na linguagem *Web Ontology Language* (OWL).

Rocha (2012) afirma que com a utilização das linguagens RDF e OWL a integração de dados passa a ser semântica e não estrutural, além disso, devem considerar que os recursos são conectados entre si, por meio dos relacionamentos expressos na ontologia (OWL) utilizando o padrão RDF, sendo desnecessária a alteração na estrutura desses recursos.

A publicação de dados utilizando os princípios do *Linked Data* pode auxiliar a representação do conhecimento. Além deste aspecto, outras vantagens da publicação de dados foram assinaladas por Grisoto (2016). Para o autor+ os conjuntos de dados publicados no *Linked Data* utilizam tecnologias da Web Semântica e padrões de metadados para descrever e representar as informações, favorecendo seu uso por pessoas e máquinas.

A possibilidade de conectar diferentes recursos por meio de links semânticos significativos tanto para usuários, quanto para máquinas, permite que grandes volumes de dados sejam acessados e manipulados de modo automático, ampliando a probabilidade de representação, busca e recuperação de dados.

As oportunidades de integração semântica dos dados na web motivam o desenvolvimento de novos tipos de aplicação e de ferramentas, como navegadores e motores de busca de acordo com Isotani e Bittencourt (2015). Essas também corroboram para ampliar as possibilidades de atuação do profissional da informação, a medida em que este pode ser um agente na construção de ontologias e vocabulários que facilitem o enriquecimento semântico das relações existente na web.

Bizer et al. (2009) afirma que a ligação entre os dados, a interoperabilidade entre sistemas e a criação de inferências, demandam prioritariamente a existência e, ou a criação de vocabulário comum entre as diferentes fontes de dados que serão conectadas. Contudo,

a maioria dos dados, amplamente disseminados na web e ou existentes em bancos de dados transacionais, fazem uso de terminologias e vocabulários fora do padrão e desconexos. O autor ressalta que a ausência de um padrão terminológico entre os vocabulários tem se tornado um grande complicador para a implementação de *Linked Data*, tendo em vista que o mesmo dado pode ser referenciado de modo distinto em diferentes vocabulários, dificultando sua reutilização e, sobretudo a sua integração. Essa afirmação pode ser confirmada por Gruber (1995) ao ressaltar que a utilização de um padrão de vocabulário comum para troca de dados permite a reutilização do conhecimento e facilita a interoperabilidade entre diferentes sistemas.

Para permitir qualificar e enriquecer semanticamente os objetos, suas representações digitais e suas diferentes relações, Marcondes (2012) relata que é necessário o desenvolvimento de um vocabulário específico – uma ontologia – de classes de objetos e processos e de relações entre estes. Para o autor as ontologias são a principal plataforma de representar a informação no *Linked Data*. (MARCONDES, 2012).

A criação de relacionamentos sem o devido cuidado ontológico, isto é, sem considerar o conjunto de relações formais, a semântica existente nessas relações e as possibilidades de criar regras de inferência, pode resultar em um vocabulário pobre e ambíguo, comprometendo a recuperação das informações e os benefícios esperados com a publicação de dados em *Linked Data*.

Mais detalhes sobre os princípios e as tecnologias usadas para a publicação de dados em *Linked Data*, são apresentados na seção a seguir.

A publicação de dados em *Linked Data* é relevante para facilitar a interoperabilidade entre sistemas e facilitar a recuperação de dados da web. Inicia-se esta seção apresentando a definição de publicação de dados.

A publicação é o processo da disponibilização permanente do conjunto de dados, por um órgão ou uma entidade, na Internet para uso irrestrito da sociedade. Este processo compreende não apenas a publicação dos dados em si, mas também dos seus metadados. (BRASIL, Cartilha 2011, p.7)

O *Linked Data* não está diretamente relacionado ao conceito de dados abertos e pode ser utilizado, inclusive, por empresas privadas que desejam disponibilizar seus dados

sem necessariamente deixá-los abertos. O Portal Brasileiro de Dados Abertos<sup>3</sup> apresenta a definição da *Open Knowledge Internacional*<sup>4</sup> para o termo. Em síntese,

Dados são abertos quando qualquer pessoa pode livremente acessá-los, utilizá-los, modificá-los e compartilhá-los para qualquer finalidade, estando sujeito a, no máximo, a exigências que visem preservar sua proveniência e sua abertura. Isso geralmente é satisfeito pela publicação dos dados em formato aberto e sob uma licença aberta.

Dados abertos conectados referem-se ao conceito denominado de LOD (Linked Open Data - Dados abertos conectados) cujos dados podem ser usados, reutilizados e redistribuídos livremente, sujeitos, no máximo, as exigências de atribuição e compartilhamento da mesma licença (ISOTANI e BITTENCOURT, 2015)

O Projeto Linking Open Data, fundado em 2007 e suportado pelo W3C, é a principal iniciativa para publicação de dados abertos conectados. O Projeto Linking Open Data tem como objetivo identificar os conjuntos de dados disponibilizados com licença aberta, para convertê-los e publicá-los na web de acordo com os princípios do *Linked Data* (BIZER; HEATH, 2011).

O diagrama da FIG. 2 The Linking Open Data Cloud Diagram apresenta o conjunto de dados abertos publicados pelo projeto Linking Open Data até o fevereiro de 2017, que traz mais de 1000 datasets (conjuntos de dados). Cada círculo (nó) representa um vocabulário criado em RDF e cada seta (arco) representa uma conexão entre os vocabulários, e as cores se referem aos domínios de conhecimento de cada vocabulário<sup>5</sup>.

Como exemplo de iniciativas Linking Open Data (LOD) - ou Ligando Dados Abertos pode ser citado o DBPedia<sup>6</sup>. Este projeto converte o conteúdo das informações

---

<sup>3</sup> Mais informações pode ser acessadas em: <http://dados.gov.br/pagina/dados-abertos>

<sup>4</sup> <https://br.okfn.org/>

<sup>5</sup> ABELE, A., MCCRAE, J. Linking open data cloud diagram. LOD Community. 2017. Disponível em: <http://lod-cloud.net/> acesso em outubro/2017

<sup>6</sup> Para maiores informações acessar: <http://dbpedia.org>

estruturadas da Wikipedia, disponibilizando na web diversos tópicos, tais como: pessoas, músicas, filmes, lugares dentre outros. É o maior conjunto de dados da web e um dos que possuem maior número de ligações com outras fontes de dados, como o GeoNames<sup>7</sup>, MusicBrains<sup>8</sup>, dentre outros.

O GeoNames é o segundo maior conjunto de dados da web e apresenta dados geográficos integrados, como os citados abaixo, mas sem se limitar a: nomes de lugares, população, altitude, latitude, longitude, mapas.

De acordo com Vandebussche et al., (2015) a iniciativa Linked Open Vocabularies (LOV) desenvolvida pela Open Knowledge Foundation é um portal para obter dados e estatísticas sobre vocabulários semânticos. Ele reúne e apresenta indicadores inéditos sobre os datasets no LOD, como a interconexão entre vocabulários, histórico de versões, política de manutenção e outros. O LOV reúne uma comunidade de cerca de 350 pessoas interessadas em vários domínios como a engenharia de ontologia ou publicação de dados.

O Gráfico 1, traz a evolução da publicação de dados em Linked Open Data, que subiu de apenas 12 datasets em maio de 2007 para 1139 em fevereiro de 2017.

---

<sup>7</sup> Para maiores informações acessar: <http://www.geonames.org>

<sup>8</sup> MusicBrainz é um banco de dados musical com conteúdo aberto. Para maiores informações acessar: <https://musicbrainz.org>

Figura 2 - The Linking Open Data cloud diagram

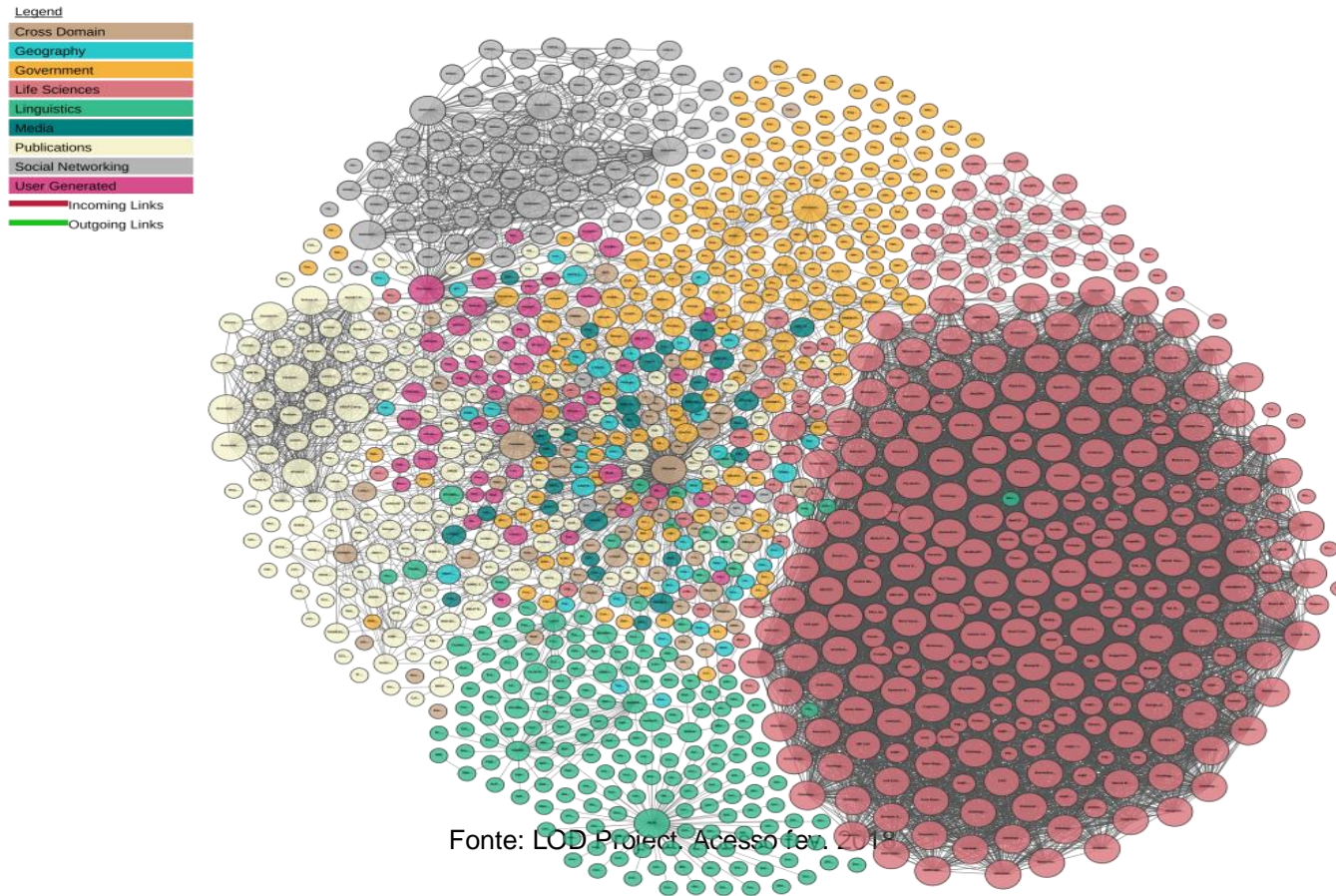
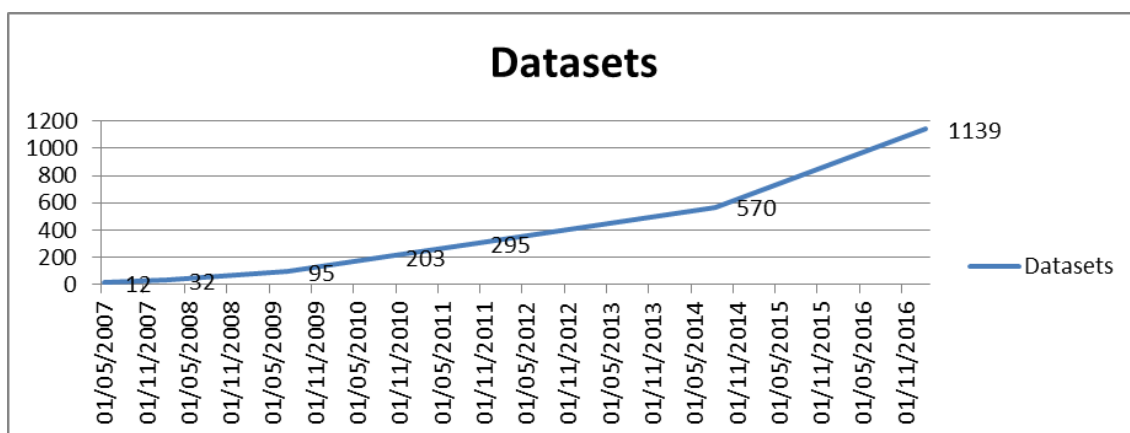


Gráfico 1 - Evolução dos Datasets



| Updated           | Datasets |
|-------------------|----------|
| <b>20/02/2017</b> | 1139     |
| <b>30/08/2014</b> | 570      |
| <b>19/09/2011</b> | 295      |
| <b>22/09/2010</b> | 203      |
| <b>14/07/2009</b> | 95       |
| <b>28/02/2008</b> | 32       |
| <b>01/05/2007</b> | 12       |

Fonte: Adaptado pelo autor do *The Linking Open Data cloud diagram*, 2017

A publicação de dados abertos no âmbito do Projeto Linked Open Data (LOD2) é composta por oito etapas denominadas ciclo de vida do LOD2, conforme apresentado na FIG.3.

Mendonça (2013) detalha as etapas do ciclo de vida do LOD2, a seguir:

1. Extração: consiste no emprego de tecnologias para acessar, extrair e triplicar dados de diferentes fontes, com o objetivo de publicá-los na web de Dados;
2. Armazenamento / Recuperação: etapa em que são utilizadas tecnologias para o armazenamento das triplas RDF, visando aumentar a escalabilidade, das consultas, do cache das junções (joins) e do processamento dos grafos;

3. Autoria / Revisão: consiste na adoção de tecnologias para facilitar a autoria de bases de conhecimento com riqueza semântica, através do paradigma "O que você vê é o que você quer dizer" (*WYSIWYM - What You See Is What You Mean*), assim como o emprego de técnicas voltadas para autoria e revisão no contexto de redes sociais distribuídas e colaboração semântica;

4. Interligação / Fusão: Utilização de tecnologias para criação e manutenção de ligações entre os dados de forma automática ou semiautomática, a fim de estabelecer coerência e facilitar a integração dos dados;

Figura 3 - Etapas do Ciclo de Vida do LOD2



Fonte: Mendonça (2013) adaptado pelo autor



5. Classificação / Enriquecimento: adoção de tecnologias para integração dos dados publicados com ontologias para permitir a integração, fusão e busca dos dados;
6. Análise de Qualidade: utilização de tecnologias para suportar a análise da qualidade dos dados, através de características como proveniência, contexto, cobertura ou estrutura;
7. Evolução / Reparo: uso de tecnologias para detectar problemas nos conjuntos de dados publicados, em virtude de mudanças ou atualizações nestes dados, vocabulários e ontologias, e automaticamente sugerir estratégias de reparo e correção;
8. Busca / Navegação / Exploração: adoção de tecnologias para busca, navegação, exploração e visualização de dados espaciais, temporais e estatísticos, publicados de acordo com os princípios de *Linked Data*, de modo a torná-los acessíveis para os usuários. (MENDONÇA, 2013, p. 35)

Auer et. al., (2013) afirmam que as etapas do ciclo de vida para publicação de dados abertos interligados previsto no Projeto LOD2 não devem ser abordadas de forma isolada. As etapas devem ser realizadas de modo integrado, permitindo a melhoria contínua do processo de publicação.

Um exemplo da necessidade de integração entre as etapas de publicação do LOD2 seria realiza em conjunto das etapas 4 (Classificação/Enriquecimento) e 7 (Evolução / Reparo). Ressalta-se que ao realizar a etapa de evolução ou reparo poderiam ser identificados ajustes na classificação/enriquecimento em virtude de problemas de modelagem. (AUER et al., 2013).

As seções a seguir apresentam os componentes necessários para a publicação de dados e a função de cada uma das tecnologias envolvidas neste processo.

### **2.2.1 Padrões e modelos para publicação em *Linked Data***

Conforme apresentado na FIG. 01 existem várias tecnologias que fazem parte da Web Semântica e do *Linked Data*. Apresentam-se a seguir sobre a URI, o Modelo RDF, OWL, SPARQL como padrões, modelos, protocolos para publicação em *Linked Data*:

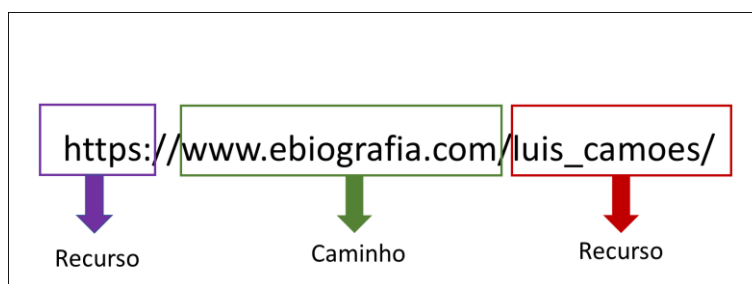
#### **2.2.1.1 URI**

Esta seção descreve o URI (*Uniform Resource Identifier*) como um dos componentes necessários para publicação de dados no *Linked Data*. Para uma melhor compreensão serão apresentados os conceitos, finalidade e estrutura do URI. Em seguida

serão detalhados os benefícios obtidos em sua utilização e as vantagens de adoção do componente URI pela BCI no processo de tratamento da informação.

O URI pode ser definido como um identificador único de recursos. De acordo com Ramalho (2006), este identificador permite a definição e adoção de nomes para estes recursos e seus respectivos endereços na internet. Trata-se de uma sequência de caracteres recomendada pelo W3C para identificar de forma única os dados na web. Esta sequência é composta por um identificador de protocolo (“http”, “ftp”, “file”, etc.) seguido pelo caractere “:” e pelo caminho para localização do recurso. (BERNERS-LEE, FIELDING, MASINTER, 2005), assim como apontado na FIG.4.

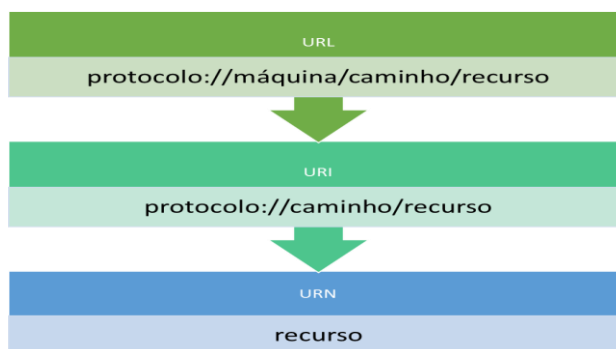
**Figura 4 - Estrutura da URI**



Fonte: Elaborado pela autora

Um URI pode ser classificado, tanto como um URN - Uniform Resource Name quanto um URL – Uniform Resource Locator. O URN - Uniform Resource Name é o identificador que define o nome de um recurso, indiferente de seu endereço ou localização). Já o URL - Uniform Resource Locator compreende o endereço de um recurso disponível em uma rede internet, intranet, rede corporativa, ou ainda a interseção destes dois identificadores, o URN e o URI. Tosin, (2016). A figura abaixo ilustra a estrutura da URI, URL e URN.

**Figura 5 - Estrutura URL, URI, URN**



Fonte: Elaborado pela autora

Os URIs podem ser utilizados para representar quaisquer objetos na web, tais como pessoas, países, músicas, filmes, entre outros, de modo que estes recursos possam ser referenciados e fornecidas outras informações sobre os mesmos. (BERNERS-LEE; FIELDING; MASINTER, 2005).

Um exemplo da utilização de URI pode ser observado ao se realizar uma pesquisa no Google sobre o apresentador Sílvio Santos. Dentre os resultados obtidos para essa busca é exibida a URI [http://dbpedia.org/resource/Silvio\\_Santos](http://dbpedia.org/resource/Silvio_Santos) que identifica o recurso Sílvio Santos no Dbpedia.

Alguns benefícios obtidos na identificação de recursos utilizando URI foram apontados por Berners-lee; Fielding; Masinter, (2005). São eles:

- 1) Permitir que diferentes identificadores de recursos sejam usados no mesmo contexto, ainda que os mecanismos para acesso a estes recursos sejam distintos;
- 2) Possibilitar interpretação semântica de modo uniforme ou padronizado de sintaxes comuns, facilitando a interoperabilidade e acesso a diferentes tipos de identificadores de recursos;
- 3) Permitir a introdução de novos tipos de identificadores de recursos sem alterar a maneira como os identificadores existentes são usados, possibilitando sua reutilização em vários e diferentes contextos.

- 4) Possibilitar que novas aplicações ou protocolos sejam utilizados para alavancar um pré-existentes ampliando a utilização de um conjunto de identificadores de recursos.

Observou-se que quaisquer recursos como vídeo, imagens, documentos, dentre outros, podem ser identificados por um URI, até mesmo os tratados em bibliotecas, instituições e unidades de informação no âmbito da BCI. O uso de URI para identificar os recursos que fazem parte do acervo informacional de uma instituição, gerando vantagens no processo de tratamento da informação poderia ser descrito como:

- a) utilização do recurso em diferentes contextos bibliotecas, museus, arquivos e outras unidades de informação;
- b) individualização do recurso, diferenciando o mesmo de qualquer outro da coleção e;
- c) reutilização do recurso por várias fontes de pesquisa sem a necessidade de duplicação.

Outra vantagem que merece destaque é que a utilização de URI (Uniform Resource Identifier) evita a redundância de informações, uma vez que um dado externo é referenciado e não replicado localmente; o que significa que cada recurso na web é identificado por um URI específico e, este por sua vez, pode ser associado a vários outros recursos, utilizando o modelo RDF sem a necessidade de duplicação (ROCHA 2012).

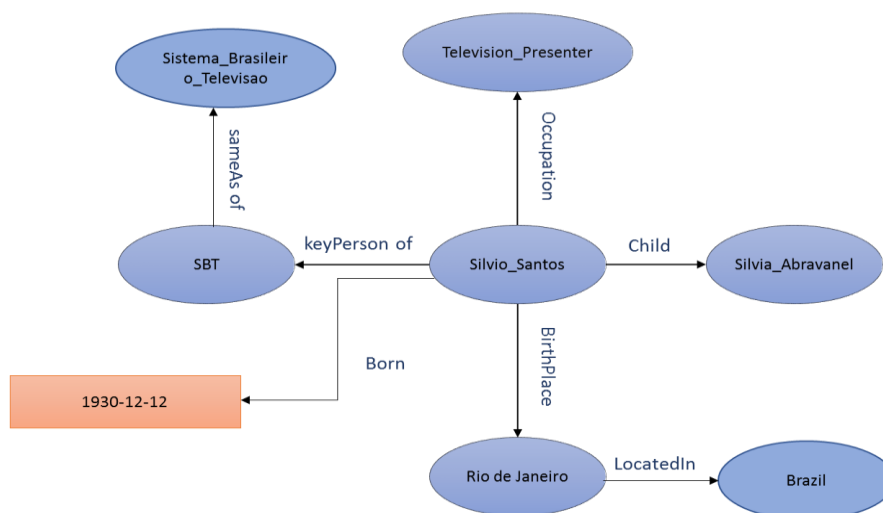
Bizer et al., (2009) também destacam a importância do uso da URI para a publicação de dados em *Linked Data*, sendo enfatizado que este é o primeiro princípio para identificar as coisas do mundo. Acrescenta-se que, nesse caso, as coisas do mundo na web podem ser expressas usando a tripla RDF (sujeito, objeto e ou predicados), modelo que será apresentado na seção a seguir.

#### 2.2.1.2 Modelo RDF

De acordo como o RDF Working Group (2014) o RDF Resource Description Framework é um modelo padronizado para troca de dados na web, sendo que possui recursos que facilitam a integração de diferentes fontes sem exigir a alteração dos dados. A sintaxe deste modelo é constituída por um conjunto de triplos, formada por um sujeito, um predicado e um objeto. Estes triplos podem estar associados a vários outros, criando

inúmeras possibilidades de busca e navegação para os usuários. Um conjunto desses triplos é chamado de grafo RDF e está representado pela FIG.6.

**Figura 6 - Exemplo de grafo RDF**



Fonte: Dbpedia adaptado pela autora.

A FIG. 6 exemplifica um grafo RDF elaborado para a consulta sobre o apresentador Sílvio Santos, considerando algumas relações desta busca na Dbpedia. As elipses desta figura representam outros recursos na internet relacionados ao apresentador, enquanto as setas mostram os relacionamentos existentes e os retângulos as literais - que são os dados sobre o recurso, a exemplo da data de nascimento ou da data de morte.

Segundo Azevedo (2014) o modelo RDF foi projetado para atender as situações nas quais a informação precisa ser processada e recuperada por aplicações, em vez de ser simplesmente visualizada por usuários. Esta situação ocorre tendo em vista que o RDF fornece um framework comum para expressar as informações, permitindo seu compartilhamento entre diferentes aplicações sem que a semântica seja perdida.

De acordo com o W3C (2014) o sujeito que corresponde a entidade do mundo real que está sendo representado. O predicado, isto é, as qualidades do sujeito sempre serão uma URI. Já o objeto pode ser uma literal (dados que representam a informação a ser descrita como: nome, número, data de nascimento, data de morte) ou uma URI para ligar a tripla a outro recurso.

Para Rocha (2014), os seguintes aspectos devem ser considerados para demonstrar a importância do modelo RDF para o *Linked Data*. São elas:

- 1) A utilização de URIs HTTP como identificadores únicos globais, possibilitando referenciar qualquer recurso;
- 2) URIs podem ser desconsideradas em um grafo RDF, durante o processo de recuperação de informações, permitindo ao usuário partir de qualquer tripla RDF como um ponto inicial de navegação de acordo com o contexto e sua necessidade de navegação;
- 3) O RDF possibilita a criação de links entre dados de diferentes fontes em quaisquer domínios, ampliando as possibilidades de pesquisa;
- 4) O RDF permite combinar informações de fontes distintas para unir dois ou mais conjuntos de triplas em mesmo grafo;
- 5) Com RDF é possível representar informações que são expressas em diferentes esquemas em um único grafo, permitindo combinar termos de vocabulários distintos para representar estes dados.

Segundo Klyne; Carroll (2004) para ser reconhecido por agentes de software o RDF possuem três níveis de abstração, sendo o sintático, o estrutural e o semântico. Estes níveis de abstração permitem que o RDF seja reconhecido por diferentes tecnologias e serão descritos da seguinte forma:

- 1) Nível sintático – Este nível trata essencialmente de documentos em XML;
- 2) Nível estrutural – Nível que apresenta o conjunto de triplas (sujeito-predicado-objeto) que codificam fatos conhecidos;
- 3) Nível semântico – Nível no qual são apresentados grafos de semântica predefinida associada a nós e arcos.

Pereira (2014) acrescenta que a capacidade do RDF de representar os metadados e seus relacionamentos semânticos em formato legível por computador, o torna um mecanismo eficiente para o intercâmbio de informação através da web, facilitando a interoperabilidade entre diversas aplicações.

O modelo RDF pode ser utilizado na BCI no processo de representação automática da informação, permitindo a interligação entre recursos provenientes de arquivos ou bibliotecas digitais. Deste modo, seria possível a integração desses recursos com outros

disponíveis na web criando grafos RDF com estes links. A conexão entre Hamlet, obra de William Shakespeare, existente em uma biblioteca, com o filme disponível na web é um exemplo. Esta conexão pode ser feita, ainda se for ligada a documentos, acerca da vida e obra do autor, que estão armazenados em arquivos ou museus. Todos estes recursos interligados ajudam a ampliar as opções de busca de um usuário mesmo que seus registros tenham sido catalogados de forma totalmente distinta.

### 2.2.1.3 OWL

De acordo com PEREIRA, (2014) a OWL (*Web Ontology Language*) é uma linguagem para publicação de ontologias recomendada pela W3C para representar o significado dos termos e seus relacionamentos em vocabulários a serem utilizados na web de dados, permitindo o processamento por agentes de software.

O conhecimento semântico proveniente de vocabulários e ontologias pode ser utilizado na publicação de dados para facilitar a busca e a recuperação dos mesmos na web. (GRUBER, 1995).

A OWL assim como o RDF é uma linguagem que permite a definição de classes, propriedades e objetos. Uma característica da OWL é que ela possui maior expressividade semântica, na medida em que possui mais construções e relacionamentos que incorporam pressupostos semântico-ontológicos, o que significa especificar com maior precisão o significado pretendido.

A reutilização de vocabulários conhecidos como FOAF<sup>9</sup> (Linguagem de computador que define um dicionário de termos relacionados as pessoas que podem ser usados em dados estruturados como RDFa e *Linked Data*), SKOS<sup>10</sup> (modelo de dados voltado para desenvolver especificações e padrões para apoiar o uso de sistemas de organização de conhecimento (KOS), como tesouros, esquemas de classificação, sistemas de cabeçalho de assunto e taxonomias no âmbito da Web Semântica) e Dublin Core<sup>11</sup>

---

<sup>9</sup> Ver mais informações em <http://www.f.oaf-project.org>

<sup>10</sup> Ver mais informações em <https://www.w3.org/2004/02/skos>

<sup>11</sup> Ver mais em [dublincore.org](http://dublincore.org)

(padrão de metadados que visa descrever objetos digitais, tais como, vídeos, sons, imagens, textos e sites na web). Para a criação de novas fontes de informação é considerada uma boa prática pelo W3C, o que também facilita o processamento dos dados pelos agentes de software. (Bizer et al., 2009).

Outra vantagem do uso da OWL é a possibilidade que ela oferece para a realização de inferências a partir dos conceitos e de seus relacionamentos, permitindo ampliar as opções de busca e a recuperação de dados.

A reutilização de ontologias é uma prática recomendada pela W3C, na medida em que facilita a interoperabilidade dos dados e amplia as possibilidades de consulta aos dados na web, principalmente utilizando a linguagem SPARQL.

A seção a seguir apresenta a linguagem SPARQL e sua importância para o *Linked Data*.

#### 2.2.1.4 SPARQL

SPARQL (SPARQL Protocol and RDF Query Language) é um protocolo e uma linguagem de consulta padrão recomendada pelo W3C para pesquisa e recuperação de informações contidas no formato RDF. O SPARQL é uma das principais tecnologias da Web Semântica e do *Linked Data*. (KENDALL; FEIGENBAUM; TORRES, 2008), (PRUD'HOMMEAUX; SEABORNE, 2008).

O SPARQL tem como objetivo realizar uma consulta semântica em bancos de dados para recuperar e manipular dados armazenados no formato RDF.

O SPARQL se assemelha ao SQL na medida em que possui uma estrutura Select-From-Where AZEVEDO (2013), onde:

- i. *Select*: Explicita uma projeção sobre os dados que serão retornados, por exemplo, a ordem e a quantidade de atributos e/ou instâncias que serão exibidas como resultado da consulta.
- ii. *From*: Informa as fontes ou bases que serão consultadas. Esta cláusula é opcional e caso não seja especificada, a busca será realizada em um documento RDF/RDFS local;
- iii. *Where*: Impõe restrições à consulta, sendo que os registros retornados pela consulta deverão satisfazer as condições impostas por esta cláusula. (Cunha et al., 2011).



A FIG. 7 apresenta a tela do sistema Virtuoso SPARQL Query Editor<sup>12</sup> utilizada para realizar uma consulta SPARQL no banco de dados da Dbpedia.

O quadro 3 demonstra um exemplo de consulta, onde o usuário deseja saber quais as entidades de língua alemã. Nesse caso, é selecionada a classe ent (entidade), e localizada na propriedade language, todas as entidades que satisfizerem a pesquisa. Traduzindo a expressão da consulta SPARQL temos:

**Quadro 3 - Transcrição da Consulta SPARQL**

| EXPRESSÃO                                     | DESCRIÇÃO  |
|---|--|
| SELECT ?ent                                   | Significa que no resultado da consulta SPARQL serão apresentadas objetos da classe <b>ent</b> (entidade do Dbpedia)  |
| WHERE (?ent dbo:language dbr:German_language) | Significa que somente serão exibidos objetos da classe <b>ent</b> , onde a propriedade linguagem ( <b>dbo:language</b> ) seja identificada pela propriedade língua alemã ( <b>dbr:German_language</b> ). |

Fonte: Elaborado pela autora

---

<sup>12</sup> <https://dbpedia.org/sparql>

Figura 7 - Tela de consulta SPARQL

D Virtuoso SPARQL Query Editor x

← → ↻ Seguro | https://dbpedia.org/sparql

### Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)  
http://dbpedia.org

Query Text  
select distinct ?Concept where {[ ] a ?Concept} LIMIT 100

(Security restrictions of this server do not allow you to retrieve remote RDF data, see [details](#).)

Results Format: HTML

Execution timeout: 30000 milliseconds (values less than 1000 are ignored)

Options:

- Strict checking of void variables
- Log debug info at the end of output (has no effect on some queries and output formats)
- Generate SPARQL compilation report (instead of executing the query)

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query Reset

Fonte: *Dbpedia* (2017)

A FIG. 8 apresenta o resultado da pesquisa na DBPedia, contendo todas os recursos que atendiam a consulta, ou seja todas as entidades de lingua alemã catalogadas no DBpedia.

Figura 8 - Resultado da consulta SPARQL



| ent   |
|---|
| <a href="http://dbpedia.org/resource/Frisians">http://dbpedia.org/resource/Frisians</a>   |
| <a href="http://dbpedia.org/resource/John_Milton">http://dbpedia.org/resource/John_Milton</a>   |
| <a href="http://dbpedia.org/resource/Kashubians">http://dbpedia.org/resource/Kashubians</a>   |
| <a href="http://dbpedia.org/resource/OS/2">http://dbpedia.org/resource/OS/2</a>   |
| <a href="http://dbpedia.org/resource/Organization_for_Security_and_Co-operation_in_Europe">http://dbpedia.org/resource/Organization_for_Security_and_Co-operation_in_Europe</a> |
| <a href="http://dbpedia.org/resource/FreeDOS">http://dbpedia.org/resource/FreeDOS</a>   |
| <a href="http://dbpedia.org/resource/A_Hunger_Artist">http://dbpedia.org/resource/A_Hunger_Artist</a>   |
| <a href="http://dbpedia.org/resource/Knoppix">http://dbpedia.org/resource/Knoppix</a>   |
| <a href="http://dbpedia.org/resource/Scilab">http://dbpedia.org/resource/Scilab</a>   |
| <a href="http://dbpedia.org/resource/Correction_(novel)">http://dbpedia.org/resource/Correction_(novel)</a>   |
| <a href="http://dbpedia.org/resource/Doctor_Crippen_(1942_film)">http://dbpedia.org/resource/Doctor_Crippen_(1942_film)</a>   |
| <a href="http://dbpedia.org/resource/Remembering_the_Kanji_and_Remembering_the_Hanzi">http://dbpedia.org/resource/Remembering_the_Kanji_and_Remembering_the_Hanzi</a>           |
| <a href="http://dbpedia.org/resource/Angst_(1928_film)">http://dbpedia.org/resource/Angst_(1928_film)</a>   |
| <a href="http://dbpedia.org/resource/Angst_(1983_film)">http://dbpedia.org/resource/Angst_(1983_film)</a>   |
| <a href="http://dbpedia.org/resource/Anni_(film)">http://dbpedia.org/resource/Anni_(film)</a>   |
| <a href="http://dbpedia.org/resource/Aviation_Safety_Network">http://dbpedia.org/resource/Aviation_Safety_Network</a>   |
| <a href="http://dbpedia.org/resource/Bomba_the_Jungle_Boy">http://dbpedia.org/resource/Bomba_the_Jungle_Boy</a>   |
| <a href="http://dbpedia.org/resource/Carmen_(1918_film)">http://dbpedia.org/resource/Carmen_(1918_film)</a>   |
| <a href="http://dbpedia.org/resource/Clarissa_(film)">http://dbpedia.org/resource/Clarissa_(film)</a>   |

Fonte: *Dbpedia*

A linguagem SPARQL permite que agentes de software possam acessar os objetos armazenados na web e identificar outros objetos relacionados, oferecendo aos usuários outros recursos na internet. No entanto, para que o SPARQL seja realizado como ferramenta de pesquisa, é fundamental a utilização de um modelo de dados que padronize o relacionamento entre os objetos.

A seção a seguir apresenta dois modelos de dados que realizam o enriquecimento semântico para a publicação de dados em *Linked Data*. São eles: *Europeana Data Model* (EDM) e *Bibliographic Framework Initiative* (BIBFRAME).

## 2.3 Modelos de Dados

Um modelo de dados pode ser definido como instrumentos de modelagem, representação e organização de objetos informacionais utilizado para favorecer a descrição formal dos recursos, e possibilitar que aplicações computacionais compartilhem e reutilizem metadados, enquanto preservam o significado original dos recursos.

Utilizar modelos de dados, dotados de forma e de padrão pode ser a resposta para representar conhecimento de determinado domínio e permitir sua recuperação. Ao mesmo tempo, tais modelos são mecanismos eficientes para representar, de forma simplificada e uniforme, a realidade a ser compreendida por usuários e máquinas, facilitando a recuperação e manipulação de grandes volumes de dados.

Para Sayão (2001) um modelo serve para comunicar algo sobre o objeto modelado, visando gerar um entendimento mais completo sobre a realidade. No entanto, para atingir este objetivo o modelador precisa possuir uma visão clara e livre de ambiguidades do objeto modelado.

Existem modelos de dados em diferentes áreas do conhecimento como medicina, agricultura, engenharia. No âmbito deste estudo foram selecionados dois modelos da área de patrimônio cultural, sendo o EDM que trabalha com acervos bibliográficos, arquivísticos e museológicos e o BIBFRAME que trabalha essencialmente com acervo bibliográfico, uma vez que são importantes referências na BCI que se propõe a garantir a interoperabilidade, integração e enriquecimento semântico de dados para sua publicação utilizando os princípios do *Linked Data*.

As seções 2.3.1 e 2.3.2 apresentam esses modelos de dados.

### 2.3.1 Europeana Data Model (EDM)

A Europeana<sup>13</sup> é uma biblioteca virtual com a maior coleção *online* de arte, cultura e ciência, incluindo coleções de bibliotecas, objetos de museus e de galerias e

---

<sup>13</sup> Ver mais informações em <http://www.europeana.eu/portal/pt/about.html>

documentos de arquivos, desenvolvida pelos países da União Europeia. Seu acervo é composto por livros e manuscritos, fotos e pinturas, esculturas e partituras musicais, vídeos e gravações de som, diários e mapas, entre outros objetos que representam o patrimônio cultural da Europa.

Para garantir a interoperabilidade entre os dados de diferentes provedores de conteúdo fornecidos pelas bibliotecas, museus e arquivos, a Europeana optou por estabelecer uma estrutura de metadados de uso comum para ser utilizada por toda comunidade fornecedora de conteúdo. (TOMITA, 2015)

Ainda segundo Tomita (2015), inicialmente a Europeana implantou o padrão de metadados *Europeana Semantic Elements* (ESE), com o objetivo de padronizar as descrições dos recursos disponibilizados para consultas. O ESE utiliza elementos oriundos do padrão de metadados Dublin Core com elementos específicos da Europeana para descrever os recursos existentes.

Em razão da diversidade de padrões adotados pelos provedores de conteúdo e de sua incompatibilidade com o ESE, que acabavam por dificultar a integração de dados e interoperabilidade entre os acervos, a Europeana propôs um novo modelo de dados denominado *Europeana Data Model* (EDM). O objetivo do EDM é 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).

Segundo Coneglian e Santarém (2017) o projeto da Europeana no LOD disponibiliza um ambiente SPARQL Endpoint, que possibilita a busca e a recuperação dos dados estruturados da Europeana, através de uma linguagem de chamada de *SPARQL Protocol and RDF Query Language* (SPARQL).

Segundo Charles e Isaac (2012) afirma que a Europeana desenvolveu o EDM 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. Ainda, segundo o autor definir um esquema de metadados cujos elementos fossem comuns a todas as comunidades de fornecedores de dados era fundamental para permitir a integração entre objetos de diferentes acervos, o que não era possível para o modelo ESE padrão de metadados adotado anteriormente pela Europeana.

Winer e Rocha (2013) aponta algumas vantagens para o uso do EDM. Segundo o autor o uso do EDM permitirá aos usuários navegar na Europeana de uma maneira mais moderna e flexível, fazendo conexões entre o objeto real e suas várias representações no ambiente digital, por meio da Web Semântica. O quadro da Monalisa no Museu do Louvre, por exemplo, pode ser relacionado às suas várias representações digitais tais como a descrição no Dbpedia, vídeos sobre a obra, fotos do acervo, dentre outros.

O uso de metadados transformados em URI, representados no padrão RDF/XML, o enriquecimento dos dados e a associação destes dados com os objetos culturais já existentes, são ações fundamentais para a conversão e a integração de dados armazenados em outras bases. De acordo com a Europeana<sup>14</sup> (2013), os requisitos que guiaram a construção do modelo EDM, são:

1. Distinção entre os objetos fornecidos (pintura, livro, filme, etc.) de suas representações digitais, ou seja, o objeto real possui identificação diferente de suas representações na web;
2. Distinção do objeto e sua gravação de metadados, ou seja, o objeto real é diferente dos dados cadastrados pelos provedores de conteúdo;
3. Possibilidade de vários registros para um mesmo objeto, ainda que contenham declarações potencialmente contraditórias sobre isso, isto é, os registros do metadados feitos pelos diferentes provedores sobre o objeto devem estar disponíveis para o usuário, mesmo que apresentem descrições ou conceitos opostos;
4. Suporte para objetos que são compostos de outros objetos, permitindo que esses sejam recuperados em conjunto, através de suas relações (todo-parte) ou individualmente;
5. Suporte para recursos contextuais, incluindo conceitos de vocabulários controlados, permitindo a reutilização de termos e relações.

Esses princípios são importantes porque ampliam a visão sobre o objeto informacional, contribuem para o enriquecimento semântico e possibilitam diferentes formas de busca e recuperação dos dados sobre o mesmo.

---

<sup>14</sup> Europeana Data Model Primer. 2013. Disponível em: < <https://goo.gl/b3xL4r> >. Acesso em: 01 ago. 2018.

Segundo Haslhofer e Isaac., (2011) o EDM adota uma estrutura baseada na Web Semântica que permite o enriquecimento dos dados a partir de uma variedade de fontes externas. Ainda segundo os autores, um objeto informacional do provedor **A** pode ser enriquecido por metadados do provedor **B**, que pode ser enriquecido pela adição de dados do provedor **C** e assim por diante, criando uma ampla teia semântica. O modelo EDM apoia esta riqueza de ligações e exibe a origem de todos os dados que conectam os recursos informacionais (Haslhofer e Isaac 2011).

Para realizar o enriquecimento semântico dos objetos disponibilizado pela Europeana, primeiramente, são inseridas anotações feitas por seu escritório, contendo informações relevantes acerca dos objetos culturais. Estes objetos são relacionados aos itens do tesaurus das representações no Tesouro GEMET.<sup>15</sup> Já nos itens referentes às pessoas são criados relacionamentos com o DBpedia<sup>16</sup> permitindo a ligação dos objetos de patrimônio cultural com URI já existentes, tornando assim os dados ligados com outras bases. (CONEGLIAN e SANTARÉM, 2017)

Os elementos do EDM são constituídos tanto por classes e propriedades próprias, quanto por elementos reutilizados de outros esquemas: *Open Archives Object Reuse and Exchange Model* (OAI-ORE), *Dublin Core* e *Simple knowledge Organization System* - SKOS. A especificação de todos os elementos se encontram descritas no documento “*Definition of the Europeana Data Model elements*” e serão apresentadas a seguir.

De acordo com Tomita (2015) a Europeana possui sete classes principais no EDM, estruturadas em três classes núcleo e quatro classes contextuais. As classes Núcleo ou “Core” permitem que o EDM realize ligações dos elementos do item real com sua representação digital através das agregações. Já as classes contextuais permitem a conexão com outras classes e provedores de conteúdo ao utilizar os recursos do *Linked Data*, permitindo a interoperabilidade e o enriquecimento semântico dos dados.

---

<sup>15</sup>General Multilingual Environmental **Thesaurus.** Mais detalhes em: <https://www.eionet.europa.eu/gemet/>

<sup>16</sup> Mais detalhes em: <http://dbpedia.org>

O quadro 4 a seguir apresenta as sete classes do EDM:Quadro 4 - Classes do EDM

| Classes                    | Nome da classe     | Descrição  |
|----------------------------|--------------------|--|
| <b>Classes núcleo</b>      | edm: ProvidedCHO – | Provedor de objetos do patrimônio cultural - Corresponde ao objeto real digitalizado ou ao objeto nativamente digital.   |
|                            | edm: webResource – | Representação Digital. Representa o objeto descrito na Europeana.  |
|                            | edm: Aggregation – | Agregação que agrupa as classes  |
| <b>Classes contextuais</b> | edm: Agent         | Quem. Classe contextual que representa pessoas ou organizações.  |
|                            | edm: Place –       | Onde. Classe contextual que representa localização espacial identificada por um provedor e denominado de acordo com algum vocabulário específico tal como o Geonames.  |
|                            | edm: TimeSpan      | Quando. Classe contextual que representa períodos de tempo.  |
|                            | skos: Concept –    | O quê. Classe contextual que refere-se a um <b>conceito</b> definido como uma unidade de pensamento ou significado advinda de uma base de conhecimento organizado, tal como um tesouro ou ontologia, onde os URI's ou identificadores locais foram definidos para representar cada conceito. |

Fonte: adaptado pela Autora de Tomita (2017) apud Europeana,

O modelo EDM da Europeana será utilizado na parte empírica deste estudo como referência, para validar a proposta metodológica para avaliar o enriquecimento semântico de dados publicados em *Linked Data* desenvolvida nessa pesquisa. As informações deste modelo servirão de fundamentação teórica para realização desta pesquisa, na medida em que agrega em sua concepção aspectos de modelagem conceitual e enriquecimento semântico para publicação de dados bibliográficos, arquivísticos e museológicos em *Linked Data*. Por esse motivo, o modelo foi considerado como um importante mecanismo para responder ao objeto desta pesquisa.

### 2.3.2 Bibliographic Framework Initiative (BIBFRAME)

Em 2011, a Biblioteca do Congresso Americano, a Library of Congress dos Estados Unidos iniciou a construção de um padrão de metadados denominado Bibliographic



Framework Initiative (BIBFRAME), com a proposta de substituir o MARC21. Este modelo de metadados foi desenvolvido para publicar os registros bibliográficos na web, utilizando os princípios do *Linked Data*, de acordo com a proposta do FRBR<sup>17</sup>. (LIBRARY OF CONGRESS, 2012), (ARAKAKI, 2016), conforme confirmado por Ramalho (2016) “o modelo BIBFRAME está baseada nas tendências e práticas contemporâneas que incentivam a *disponibilização e interligação de dados na web em formatos que sejam legíveis por máquinas*”.

De acordo com a Library of Congress<sup>18</sup> (2012) o BIBFRAME tem por objetivo implementar um novo ambiente para as bibliotecas na web, que favoreça a interconexão de informações e o intercâmbio de dados.

De acordo com a Libray of Congress (2012) o BIBIFRAME é mais do que uma mera substituição do MARC 21. O modelo representa um alicerce para o futuro da descrição bibliográfica na web. Ainda segundo a instituição, o BIBIFRAME realizará seus objetivos, por meio das seguintes recomendações:

1. Diferenciação clara entre o conteúdo e suas manifestações físicas, ou seja, distinção entre o objeto real e suas, fotos, imagens, vídeos e demais representações do mesmo;
2. Identificação inequívoca de entidades de informação, por exemplo, autoridades,
3. Exposição dos relacionamentos entre as entidades.

Similar ao EDM, o BIBFRAME é composto por um conjunto de elementos e atributos estruturados em **classes**, que identificam um tipo de recurso (obra, instância, autoridade ou anotação), e **propriedades**, meios para descrever um recurso (título da obra, publicador, data de publicação, etc.). (SILVA et. al., 2017). Apresenta-se, no próximo quadro a descrição das classes principais do BIBFRAME:

---

<sup>17</sup> *Functional Requirements for Bibliographic Records* - Requisitos Funcionais para Registros Bibliográficos

<sup>18</sup> <http://www.loc.gov/bibframe/>

**Quadro 5 - Classes do BIBFRAME**

| <b>Classe</b>                 | <b>Descrição</b>   |
|-------------------------------|--|
| <b>Work (Obra Criativa) -</b> | Identificação da essência conceitual do item catalogado.   |
| <b>Instance (Instância)</b>   | Recurso que reflete a personificação material e individual da obra.  |
| <b>Authority (Autoridade)</b> | Recurso que reflete conceitos-chave de autoridade que definem relações refletidas na obra e na instância, incluindo pessoas, lugares, tópicos, organizações. |

Fonte: ARAKAKI (2016)

Ramalho (2016) afirma que além das três classes principais o BIBFRAME apresenta dezenas de outras classes e subclasses (autoridade, conceitos; lugares, dentre outros) utilizadas para categorizar diferentes tipos de recursos.

Ainda, segundo o autor o BIBIFRAME é baseado na utilização de links e identificadores controlados, que favorecem a reutilização de recursos de autoridades - diminuindo os custos associados à manutenção desses e a redução de retrabalho e de duplicidade de informação. Além disto, a utilização de links resulta em uma maior visibilidade dos recursos na web, extrapolando os limites tradicionais dos catálogos das bibliotecas.

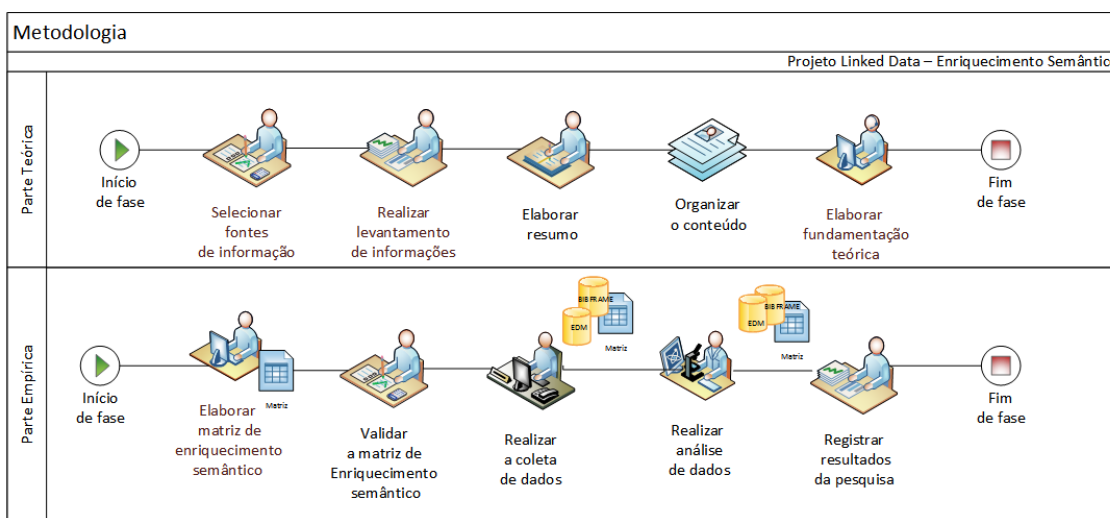
De acordo com Silva et. al., (2017) o BIBFRAME, depende fortemente das relações entre recursos (relações entre duas ou mais obras; entre obras e instâncias; entre obras e autoridade), uma vez que sua estrutura utiliza o RDF para conectar os metadados do modelo. São justamente estas relações que permitem acrescentar significado e enriquecer semanticamente os registros das obras na web, ampliando as possibilidades de busca e recuperação.

Além do modelo EDM, já apresentado anteriormente o modelo BIBFRAME aqui descrito também foi utilizado na etapa empírica desta pesquisa, para validar a proposta metodológica para avaliar o enriquecimento semântico de objetos para a publicação de dados em *Linked Data*. Os procedimentos metodológicos da pesquisa serão descritos no próximo capítulo.

### 3 METODOLOGIA

As seções seguintes descrevem as etapas desta pesquisa também ilustrada na FIG.9 a seguir:

Figura 9 – Metodologia



Fonte: Elaborado pela autora

A estrutura proposta para a metodologia (FIG. 9) é dividida em duas partes: teórica e empírica. A parte teórica apresenta os principais passos executados para a obtenção do conhecimento necessário para a elaboração dos fundamentos teórico metodológicos da pesquisa. A parte empírica apresenta a etapas adotadas para atingir os objetivos da pesquisa detalhados no capítulo 4.

#### 3.1 Caracterização da Pesquisa

A classificação da metodologia de pesquisa adotada reflete os critérios apresentados por Gil (1994), Lakatos e Marconi (1991) quanto à natureza da pesquisa, aos objetivos, pela abordagem do problema e quanto aos procedimentos técnicos, conforme apresentado a seguir:

I. **Natureza da pesquisa:** pesquisa aplicada, na medida em que visa gerar conhecimento para aplicação prática voltada na solução de problemas específicos.

II. **Objetivos:** pesquisa exploratória, haja vista que visa proporcionar maior familiaridade com um problema.

III. **Forma de abordagem do problema:** apresenta um viés qualitativo, na medida em que considera a relação existente entre o mundo e o sujeito, que não pode ser traduzida em unicamente em números.

### 3.2 Pesquisa bibliográfica

A pesquisa bibliográfica foi realizada consultando as seguintes fontes de informação: Banco de Teses e Dissertações Sucupira; *Journal of Knowledge Management*, *Knowledge Management Research Practice*, *Perspectivas em Ciência da Informação*; *Journal of the American Society for Information Science and Technology*; Scielo; Enancib; *Annual Review of Information Science and Technology*; *Journal of Knowledge Organization/ISKO\_KO* utilizando os termos “ *Linked Data*”; “ *Linked Open Data*”; “modelagem conceitual”; “modelos semânticos”; “enriquecimento semântico”, “representação do conhecimento”; “*knowledge organization*”; “*ontology*”. O levantamento bibliográfico foi realizado no período de abril/2016 a agosto/2017, considerando os últimos cinco anos de trabalhos publicados em periódicos nacionais e internacionais em português, inglês e espanhol.

Tal levantamento foi realizado com o objetivo de conhecer os conceitos que foram abrangidos na pesquisa e para estabelecer os critérios de identificação, sistematização, tecnologias envolvidas e princípios necessários para o enriquecimento semântico de dados para sua publicação em *Linked Data*.

### 3.3 Subsídios teóricos para a construção da matriz proposta

Essa etapa teve como objetivo o desenvolvimento da base teórica para a discussão dos diferentes pontos de vista identificados na literatura sobre o tema. Esta etapa possibilitou a criação de subsídios para a elaboração da fundamentação teórica desta

dissertação, bem como a construção da Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos – TRESO que será detalhada no próximo capítulo.

Abaixo são apresentadas as bases teóricas para a construção Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos – TRESO.:

1) MOURA e CAVALCANTE (2009) apresenta a anotação semântica de documentos para seu armazenamento em um SGBD (Sistema de Gerenciamento de Banco de Dados), baseado em uma ontologia de domínio, com o objetivo de melhorar o desempenho das consultas.

2) Pita e Paixão (2010) descrevem o uso de agentes para compartilhamento de ontologias, onde as necessidades dos usuários são repassadas para os agentes de software que se encarregarão de enriquecer semanticamente a requisição original, utilizando inferência lógica e outras técnicas de inteligência artificial para descobrir outros recursos da web para atender a requisição.

3) Silva (2014) apresenta vários modelos de anotação semântica, incluindo *tags*, atributos, relações e ontologias, onde são apresentadas as limitações e benefícios da utilização de cada modelo de anotação, sob o prisma do formalismo e riqueza estrutural e, da exigência requerida do usuário envolvido no processo de anotação.

4) Silva (2016) apresenta um levantamento bibliográfico sobre Anotação Semântica, abrangendo conceitos, componentes, ferramentas, tecnologias e suas aplicações, com o objetivo de auxiliar na implementação do processo de enriquecimento semântico.

5) Lira (2014) apresenta a descrição do processo de enriquecimento semântico, os atributos adotados para auxiliar na composição de metadados enriquecidos, definições e utilização de anotações semânticas, a finalidade do repositório de metadados no processo de Enriquecimento Semântico e exemplos do uso de metadados enriquecidos.

6) Sorrentino et al. (2013) apresentam o conceito de vinculação e mapeamento de recursos para interligação de dados e metadados com recursos na nuvem LOD.

### 3.4 Delimitação da Pesquisa

Na seção 2.3 foram apresentados os dois modelos de dados que realizam o enriquecimento semântico para publicar seus dados advindos de diferentes fontes na web, a saber:

1. *Europeana Data Model* (EDM)
2. *Bibliographic Framework Initiative* (BIBFRAME).

Tais modelos foram selecionados para compor a amostra desta pesquisa e foram utilizados na etapa de coleta de dados. Estes modelos de dados foram escolhidos porque ambos utilizam técnicas e padrões de modelagem conceitual e o enriquecimento semântico para publicação de dados em *Linked Data*, foco deste estudo.

Ressalta-se que esses modelos de dados apresentam diferentes estruturas, classes, propriedades e recomendações distintas para esta publicação de dados. Por este motivo é possível compará-los e identificar os pontos relevantes que podem ser aplicados a outros catálogos, ou quaisquer acervos para a publicação na web.

### 3.5 Procedimentos Metodológicos

**Procedimentos técnicos:** foi adotado como procedimento técnico para obtenção dos dados uma associação de pesquisa bibliográfica e documental. Para a construção da metodologia, foi necessário identificar os principais requisitos e técnicas de enriquecimento semântico de objetos para publicação de dados em *Linked Data* na literatura. Para a construção e a validação da Matriz proposta, utilizou-se 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. O capítulo 4 apresenta a metodologia proposta para a construção da Matriz TRESO.

## 4 CONSTRUÇÃO DA MATRIZ DE TÉCNICAS E RECURSOS PARA O ENRIQUECIMENTO SEMÂNTICO DE OBJETOS

No capítulo 2 foram apresentados, na visão de diferentes autores os conceitos, os recursos e as técnicas de enriquecimento semântico que possibilitam tornar os dados mais qualificados a partir da atribuição de significados a esses dados. Para atingir o objetivo principal deste estudo percebeu-se a necessidade de verificar quais os recursos e as técnicas são requeridos para a implementação do enriquecimento semântico. **A Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos** foi construída para possibilitar a verificação de 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 Moura e Cavalcante (2009); Lira (2014); Silva (2016); Silva (2014); Pita e Paixão (2010) e Sorrentino et al., (2013).

O Quadro 6, a seguir corresponde à **Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos**, aqui chamada apenas de Matriz TRESO.

Essa matriz foi utilizada na etapa de coleta de dados, com o objetivo de identificar os recursos e técnicas presentes em cada um dos modelos pesquisados. De posse destes dados foi feita uma análise para comparar o nível do enriquecimento semântico dos modelos de dados selecionados e os recursos e técnicas identificados na literatura.

## 4.1 4.1 CRITÉRIO DE AVALIAÇÃO

Na construção da Matriz TRESO foram criados os seguintes elementos como critérios de avaliação.

1. Anotação por tags em linguagem natural. Neste critério é avaliado se o modelo de dados utiliza a anotação por tags para fazer o enriquecimento de objetos publicados em *Linked Data*, e em quantas classes do modelo este tipo de anotação está presente.

2. Anotação semântica utilizando atributos/propriedades (proveniência, pessoa, local): Neste critério é avaliado se o modelo de dados utiliza a anotação por atributos ou propriedades para realizar o enriquecimento de objetos publicados em *Linked Data*, e em quantas propriedades do modelo este tipo de anotação está presente.

3. Anotação por meio de relações, interligando recursos através de links (relações): Este critério avalia se o modelo de dados permite a inclusão de anotações de relacionamento para enriquecimento de objetos publicados em *Linked Data*, utilizando exclusivamente as relações estabelecidas pelo modelo, ou se também adota relações estabelecidas em outros padrões.

4. Anotação semântica utilizando elementos do vocabulário controlado: Este critério avalia se o modelo de dados permite a anotação semântica para enriquecimento de objetos publicados em *Linked Data*, utilizando elementos de um único vocabulário controlado em algumas de suas classes ou em todas as suas classes, adotando diferentes vocabulários.

5. Anotação semântica utilizando ontologia: Este critério avalia se o modelo de dados possibilita a anotação semântica adotando ontologias, para fazer o enriquecimento de objetos para sua publicados em *Linked Data*, de modo a identificar formalmente conceitos dos objetos e suas relações. Para obter pontuação neste critério será avaliado se o modelo utiliza a anotação em alguma classe específica ou em todas as suas classes; ou se permite a adoção de diferentes ontologias ou de uma única ontologia.



6. Reuso de metadados: Este critério avalia se o modelo de dados realiza a reutilização de dados e metadados na publicação de dados em *Linked Data*, otimizando o trabalho do publicador e se esta reutilização ocorre em alguma classe específica, ou em todas as classes do modelo.

7. Links entre os dados e metadados do modelo com outros recursos da web: Este critério avalia se o modelo de dados realiza combinações semânticas de seus dados e metadados com alguns datasets específicos ou com quaisquer datasets da web de dados, como o Geonames e a Dbpedia.

8. Modelagem de dados num formato semântico estruturado: Este critério avalia se o formato de estruturação semântica do modelo é realizado de forma padronizada, permitindo sua manipulação por aplicações que consomem esses modelos de dados, ampliando as possibilidades de conexão com outros Datasets do LOD.

9. Utilização de ferramentas para o enriquecimento semântico: Este critério avalia se o modelo adota alguma ferramenta tecnológica para facilitar o processo de enriquecimento semântico de modo automático ou semiautomático.

10. Utilização de interface gráfica: Este critério avalia se o modelo adota alguma interface gráfica para apoiar a execução do processo de enriquecimento semântico e publicação de dados de modo automático ou semiautomático.

11. 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.

12. Agregação de metadados de diferentes fontes e vocabulários: Este critério avalia se o modelo realiza a inclusão de dados e metadados de uma única fonte ou de diferentes fontes e vocabulários, permitindo a interoperabilidade de dados.

Os critérios foram organizados na Matriz, e inicialmente foram estabelecidos os parâmetros para pontuação, considerando uma escala de 0 (zero) a 2 (dois), em que cada modelo de dados seria pontuado de acordo com a forma de atendimento de cada critério, conforme descrito no Quadro 6.

Quadro 6 - Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos – versão 1.0

(Continua)

| CRITÉRIO DE AVALIAÇÃO   | INEXISTENTE  | PARCIAL   | TOTAL   |
|---|--|---|---|
| <b>Pontuação</b>  | <b>0</b>   | <b>1</b>  | <b>2</b>  |
| 1. Anotação por <i>tags</i> em linguagem natural.                                     | O modelo não permite a realização de anotação semântica <i>utilizando tags</i> .                           | O modelo de dados permite a inclusão de anotações em algumas classes  | O modelo dados permite a inclusão de anotações e <i>tags</i> em quaisquer classes do Modelo   |
| 2. Anotação semântica utilizando atributos/propriedades (proveniência, pessoa, local) | O modelo de dados não permite a realização de anotação semântica de atributos/propriedades                 | O modelo de dados permite a inclusão de anotações de alguns dos atributos do modelo, ou em algumas de suas classes.                   | O modelo dados permite a inclusão de anotações de todas as propriedades do modelo, ou em algumas propriedades   |
| 3. Anotação por meio de relações.   | O modelo de dados não permite a quaisquer anotações para inclusão de links ou relações entre os conceitos. | O modelo de dados permite a inclusão de anotações de relacionamento, utilizando exclusivamente as relações estabelecidas pelo modelo. | O modelo de dados permite a inclusão de anotações de relacionamento, utilizando tanto as relações estabelecidas pelo modelo, quanto relações estabelecidas em outros padrões. |

(continuação)

| CRITÉRIO DE AVALIAÇÃO   | INEXISTENTE   | PARCIAL  | TOTAL  |
|---|---|--|--|
| 4. Anotação semântica utilizando elementos do vocabulário controlado.     | O modelo de dados não permite anotações utilizando elementos de vocabulário controlado.     | O modelo de dados permite a anotação semântica, utilizando elementos de um único vocabulário controlado ou em algumas de suas classes. | O modelo de dados permite a inclusão de anotações semânticas, utilizando diferentes vocabulários em todas as classes.                                  |
| 5. Anotação semântica utilizando a ontologia .                            | O modelo de dados não permite anotação semântica utilizando ontologias.                     | O modelo de dados possibilita a anotação semântica de ontologias em algumas classes específicas.                                       | O modelo de dados permite adoção de modelos semânticos, utilizando diferentes ontologias ou a mesma ontologia em todas as suas classes.                |
| 6. Reuso de metadados.  | 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 algumas classes específicas.  | O modelo de dados permite a reutilização de diferentes metadados em todas as classes.  |
| 7. Links entre os dados e metadados do modelo 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 alguns <i>datasets</i> específicos                       | 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. |

(continuação)

| CRITÉRIO DE AVALIAÇÃO  | INEXISTENTE  | PARCIAL   | TOTAL   |
|--|--|---|---|
| 8. Modelagem de dados num formato semântico estruturado.                                   | O modelo de dados não prevê a utilização de formato semântico estruturado, impossibilitando sua manipulação por agentes de software. | Em algumas classes 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 |
| 9. Utilização de ferramentas para o enriquecimento semântico.                              | O modelo realiza o enriquecimento semântico de forma manual.   | O enriquecimento semântico é realizado de forma semiautomática.   | Todo o processo de enriquecimento semântico é feito de forma automática, utilizando ferramentas para realizar anotações ou vinculação de dados.           |
| 10. Utilização de interface gráfica para o enriquecimento semântico e publicação de dados. | Inexistência de interface gráfica para facilitar o enriquecimento semântico pelo publicador.   | 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.                                     |
| 11. Relações entre os metadados do modelo e outros vocabulários.                           | O sistema não realiza controle de sinonímia, permitindo a inclusão de termos similares ou duplicados.                                | O modelo estabelece relações de sinonímia (equivalência) com alguns termos, mas em algumas classes existem termos ambíguos ou similares | Utilização de relações de sinonímia (equivalência) entre o metadado e o termo correspondente em outros vocabulários utilizados, reduzindo a ambiguidade.  |

| <b>(conclusão)</b>  |   |  |   |
|---|---|--|---|
| <b>CRITÉRIO DE AVALIAÇÃO</b>                                    | <b>INEXISTENTE</b>                                      | <b>PARCIAL</b>   | <b>TOTAL</b>  |
| 12. Agregação de metadados de diferentes fontes e vocabulários. | O modelo de dados não realiza a agregação de metadados. | O modelo de dados realiza a agregação de dados e metadados de uma única fonte. | O modelo de dados realiza a agregação de dados e metadados de diferentes fontes e vocabulários. |

Fonte: Elaborado pela autora

Como instrumento de coleta de dados para aplicação da Matriz TRESO nos modelos EDM e BIBFRAME foi utilizado o Formulário de Avaliação de Enriquecimento semântico, Quadro 8, apresentado a seguir. A primeira coluna foi utilizada para descrever o critério avaliado de acordo com a Matriz TRESO; a segunda coluna foi utilizada para inserir a pontuação atribuída a cada modelo de acordo com o critério de avaliação definido na matriz, Quadros 6 e a terceira coluna para descrever as justificativas para a pontuação atribuída, após análise e realizado à luz da literatura e dos parâmetros de pontuação da matriz TRESO.

**Quadro 7 - Formulário de Avaliação de Enriquecimento semântico**

| <b>Critério de Avaliação</b> | <b>Pontuação</b>                  | <b>Justificativa</b>                     |
|------------------------------|-----------------------------------|--|
| 1. Critério avaliado         | Pontuação atribuída a cada modelo | Justificativa para a pontuação atribuída |
| 2.                           |                                   |  |
| 3.                           |                                   |  |
| 4.                           |                                   |  |
| 5.                           |                                   |  |
| 6.                           |                                   |  |
| 7.                           |                                   |  |
| 8.                           |                                   |  |
| 9.                           |                                   |  |
| 10.                          |                                   |  |
| 11.                          |                                   |  |
| 12.                          |                                   |  |

Fonte: elaborado pela autora

O instrumento de coleta de dados foi preenchido com o registro da pontuação de cada critério, bem como com o registro da justificativa para o recebimento da nota alcançada.

O formulário de Avaliação de Enriquecimento Semântico foi aplicado a cada um dos modelos, visando identificar se os modelos de dados EDM e BIBFRAME atendem ao processo de enriquecimento semântico para a publicação de objetos em *Linked Data*.

Ressalta-se que antes de aplicar a Matriz TRESO proposta neste, foi realizada uma etapa de validação para analisar a aplicação do instrumento proposto, conforme apresentado na seção 5.1 do capítulo a seguir.

## 5 RESULTADOS E ANÁLISE

No presente capítulo foram apresentados e discutidos os resultados da coleta de dados realizada para a pesquisa, a partir da utilização da metodologia proposta no capítulo três e da matriz criada no capítulo 4. A seção 5.1 apresenta o resultado da validação da matriz, por meio de sua aplicação em cada um dos modelos e a versão final da Matriz TRESO versão 2.0 após a validação. A seção 5.2 abrange os procedimentos aplicados ao modelo de dados EDM. E, por último a seção 5.3, compara os resultados obtidos em cada modelo.

### 5.1 Validação da matriz

Após a construção da matriz, o formulário de avaliação semântica resultante foi testado de forma a verificar se os critérios definidos na matriz poderiam ser aplicados aos Modelos de Dados selecionados. Para tanto, foi escolhido um dos critérios para ser usado na etapa de validação da matriz, por meio da aplicação do formulário nos modelos EDM e BIBFRAME.

A validação de dados foi feita na amostra da pesquisa, devido à ausência de outros modelos com as mesmas características desta amostra. Assim, para a execução desta etapa realizada a validação apenas em uma parte do modelo, utilizando o **critério de avaliação nº 01 - Anotação semântica, utilizando tags em linguagem natural**.

A seguir serão apresentados os resultados da validação dos instrumentos que foram utilizados na etapa empírica dessa pesquisa e que foram implementados nos dois modelos de dados estudados, bem como a matriz atualizada após a revisão.

#### 5.1.1 Validação da matriz no Modelo BIBFRAME

Inicialmente, foi analisada a documentação do BIBFRAME disponibilizada pela *Library of Congress*, para verificar se o modelo permitia a anotação semântica por meio das *tags*. Após esta constatação foi averiguada quais as classes do modelo essas anotações foram realizadas. Na sequência, foi identificado o grau de atendimento ao critério analisado,



depois atribuída a pontuação pertinente. Por último, e após todas as verificações anteriormente citadas, foi feito o registro das justificativas para a pontuação de cada item avaliado, conforme apresentado no quadro a seguir:

**Quadro 8 - Validação da Matriz no BIBFRAME**

| CRITÉRIO DE AVALIAÇÃO   | PONTUAÇÃO | JUSTIFICATIVA  |
|---|-----------|--|
| 1. Anotação semântica utilizando <i>tags</i> em linguagem natural | 1         | <p>No BIBFRAME o recurso de anotação é utilizado para fornecer informações adicionais a outros recursos do modelo. <i>Library of Congress</i> (2012). Apesar de suportar anotação semântica aumentando o conhecimento sobre outro recurso por meio da inclusão de <i>tags</i> no BIBFRAME este recurso está relacionado a uma única classe específica <i>Annotation que pode ser associada as propriedades</i>:</p> <ul style="list-style-type: none"> <li>• <i>BF: annotates</i>: Recurso ao qual a anotação se refere</li> <li>• <i>BF: annotationAssertedBy</i>: Anotação foi afirmada pela entidade dada.</li> <li>• <i>BF: annotationBody</i>: Teor da anotação sobre o recurso.</li> <li>• <i>BF: annotationSource</i>: Fonte da anotação.</li> <li>• <i>BF: hasAnnotation</i>: <i>Resource</i> tem uma anotação.</li> </ul> <p>Deste modo, trata-se de um atendimento parcial neste item, sendo, portanto, estabelecida a pontuação 1, por ser referente a uma única classe e não a todas as classes do modelo.</p> |

Fonte: Elaborado pela autora

### 5.1.2 Validação da matriz no Modelo EDM

Da mesma forma como foi procedida a análise do critério 01 no Modelo BIBFRAME foi aplicada no EDM, disponibilizado pela Europeana para verificar como o modelo realiza a anotação semântica por meio das *tags*. O quadro 9 a seguir apresenta o resultado desta análise.

**Quadro 9 - Validação da Matriz no EDM**

| Critério de Avaliação   | Pontuação | Justificativa  |
|---|-----------|--|
| 1. Anotação semântica utilizando <i>tags</i> em linguagem natural | 1         | <p>O EDM apresenta o recurso de anotação, utilizando rotulagem simples(tags) e semântica e vinculação de objetos.</p> <p>O modelo de anotações é baseado no padrão do modelo de dados da anotação da web do W3C.</p> <p>O objeto da anotação representa os recursos com o qual a anotação se refere. No contexto da Europeana, uma classe que normalmente recebe anotações é um objeto de patrimônio cultural fornecido (edm: ProvidedCHO) ou um recurso da web associado a ele, utilizando tags (literais) ou referências a vocabulários externos utilizando o rdf.</p> <p>Assim como no Bibframe a anotação está vinculada a algumas classes e propriedades, sendo EDM pontuado neste critério também com a nota 1 – atendimento parcial, tendo em vista que existem classes no EDM onde não é realizada a anotação por <i>tags</i>.</p> |

Fonte: Elaborado pela autora

A validação realizada permitiu verificar o potencial da Matriz TRESO e os critérios estabelecidos para a avaliação dos modelos de dados. Após esta etapa, percebeu-se a necessidade de alguns ajustes na Matriz, para otimizar a análise do potencial do enriquecimento semântico para a publicação em *Linked Data*.

Durante a etapa de validação, quando foi aplicado o critério 01 **Anotação por tags em linguagem natural, para enriquecimento de objetos publicados em *Linked Data***, da Matriz TRESO foram observados os seguintes pontos:

a) A literatura citada no referencial teórico desse estudo, principalmente o trabalho sobre os tipos de anotação semântica de Silva (2014), demonstrou que o padrão de anotação empregado e a possibilidade do uso deste para conectar dados de diferentes fontes no âmbito do *Linked Data* é mais relevante para o enriquecimento semântico, do que avaliar em quantas e em quais classes o recurso de anotação está sendo utilizado.

b) A forma de pontuação do critério, previsto anteriormente na Matriz TRESO, se mostrou inadequada para diferenciar o modo com que cada modelo de dados realiza a anotação semântica; na medida em que considerava a quantidade de classes em que o recurso era apresentado, e não as técnicas de anotação realizada.

Em virtude das considerações descritas anteriormente e identificadas na etapa de validação, optou-se por agrupar os requisitos de diferentes tipos de anotação da Matriz TRESO. Da mesma forma, atribuiu-se maior pontuação a estes requisitos, de acordo as potencialidades que cada anotação representa no processo de enriquecimento semântico.

Tal agrupamento foi realizado porque durante o processo de reavaliação da Matriz TRESO e do estabelecimento dos parâmetros de pontuação, foi possível observar que o comportamento nos dois modelos era idêntico. Após nova análise da literatura constatou-se que os requisitos e técnicas eram bastante similares, o que resultou na junção destes dois critérios.

Outro ponto onde foram necessários ajustes foi na escala de pontuação definida para cada modelo dentro dos critérios da Matriz TRESO. Na escala anterior cuja pontuação variava de zero (0) a dois (2) mesmo se um modelo apresentasse uma performance superior a outro modelo suas pontuações poderiam ser bastante parecidas em virtude. Deste modo, para cada critério avaliado, foi atribuído ao Modelo de Dados a pontuação de zero (0) a quatro (4), de acordo com o grau de atendimento de cada modelo, considerando a escala de pontuação descrita no quadro 10:

**Quadro 10 - Escala de Pontuação por critério**

| <b>Pontuação</b> | <b>Descrição</b>   |
|------------------|--|
| 0 (zero)         | Será atribuída a pontuação zero para o modelo que não apresentar o critério apresentado. |
| 1 (um)           | Modelo que apresenta alguns dos recursos propostos                                       |
| 2 (dois)         | Modelos que apresentem até dois dos requisitos de um critério.                           |
| 3 (três)         | Mais de dois requisitos de um critério   |
| 4 (quatro)       | Atendem integralmente a todos os requisitos propostos pelo critério.                     |

Fonte: Elaborada pela autora

O quadro 11 a seguir apresenta a escala de pontuação que pode ser obtida pelos modelos após a aplicação de todos os critérios propostos pela Matriz TRESO.

Quadro 11 - Escala de Pontuação Total da Matriz

| Pontuação | Grau de Atendimento      | Descrição  |
|-----------|--------------------------|--|
| 0         | Não atende               | Serão classificados com esta pontuação os modelos que não atenderem a nenhum critério da Matriz TRESO, sendo aqueles modelos que não aplicam nenhuma das técnicas propostas pela matriz. |
| 1 a 7     | Atendimento insuficiente | Serão classificados com grau de atendimento insuficiente os modelos que apresentam baixa aderência aos critérios propostos na matriz TRESO.  |
| 8 a 14    | Atendimento parcial      | Serão classificados com grau de atendimento parcial os modelos que atenderem de forma moderada aos critérios propostos na Matriz TRESO.<br><b>Continua</b>                               |
| 15 a 21   | Atendimento satisfatório | Serão classificados com grau de atendimento satisfatório os modelos que tenham uma boa avaliação em grande parte dos critérios propostos.  |
| 22 a 28   | Atendimento Total        | Serão classificados com grau de atendimento total os modelos que atendam com excelência a maioria dos critérios.   |

Fonte: Elaborado pela autora

No quadro 12 a seguir, é apresentada a versão 2.0 da Matriz TRESO ajustada a partir da etapa de validação. Ressalta-se, ainda que nesta nova versão da Matriz TRESO o número de critérios de avaliação passou por uma etapa de agrupamento e foi reduzido de 12 para 7. Vale lembrar que a versão 2.0 da Matriz TRESO foi implementada neste estudo.

Quadro 12 - Matriz TRESO versão 2.0

(Continua)

| Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos  |  | Modelo de dados:(    ) EDM (    )BIBFRAME<br>Data: ____/____/____   |  |   |   |
|---|--|---|--|---|---|
| CRITÉRIO DE AVALIAÇÃO   | INEXISTENTE  | EM DESENVOLVIMENTO  | BAIXO  | ATENDE  | ATENDIMENTO TOTAL   |
| Pontuação   | <b>0</b>   | <b>1</b>  | <b>2</b>   | <b>3</b>  | <b>4</b>  |
| 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> , |

| 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   | <b>0</b>  | <b>1</b>   | <b>2</b>   | <b>3</b>   | <b>4</b>   |
| 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 específicas 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. |

(Continuação)

| Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos                   |  |  | Modelo de dados:(    ) EDM (    )BIBFRAME<br>Data: ____/____/____  |  |  |
|--|--|--|--|--|--|
| CRITÉRIO DE AVALIAÇÃO  | INEXISTENTE  | EM DESENVOLVIMENTO   | BAIXO  | ATENDE   | ATENDIMENTO TOTAL  |
| <b>Pontuação</b>   | <b>0</b>   | <b>1</b>   | <b>2</b>   | <b>3</b>   | <b>4</b>   |
| 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                    | O modelo de dados não prevê a utilização de formato semântico estruturado,   | Em apenas uma classe do modelo são utilizados formatos semânticos e estruturados como o              | Em apenas até duas classe do modelo são utilizados formatos semânticos e                                 | Os dados são modelados num formato semântico estruturado, como o   | Os dados são modelados num formato semântico estruturado, como o RDFS e OWL beneficiando sua   |



| 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  |
| manipulação por aplicações que consomem esses modelos de dados, ampliando as possibilidades de conexão com outros Datasets do LOD. | impossibilitando sua manipulação por agentes de software.    | RDF/XML.  | estruturados como o RDF/XML.  | RDF/XML, beneficiando sua manipulação por aplicações que consomem esses modelos de dados  | 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 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  | Inexistência de interface gráfica para facilitar             | A interface gráfica é utilizada para uma única função, sendo o                                  | A interface gráfica é utilizada para algumas funções,                                       | Os publicadores de dados utilizam a interface gráfica para  | Os publicadores de dados e usuários da informação utilizam a interface gráfica   |

| 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  |
| de enriquecimento semântico e publicação de dados.  | enriquecimento semântico pelo publicador.  | processo de enriquecimento semântico semiautomático.   | sendo o processo de enriquecimento semântico semiautomático.                            | apoiar a execução do processo de enriquecimento semântico.  | 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 pela autora

## 5.2 Estudo comparativo entre os modelos de dados EDM e BIBFRAME

Nesta etapa, a proposta foi identificar se os modelos EDM e BIBFRAME atendiam aos recursos e técnicas para o enriquecimento semântico listados na matriz proposta no quadro 12.

Para tanto, cada modelo foi avaliado de acordo com os critérios definidos na **Matriz de Técnicas e Recursos para o Enriquecimento Semântico de Objetos**, utilizando o Formulário de Avaliação de Enriquecimento Semântico descrito anteriormente.

Após a pontuação do modelo em cada critério o resultado foi somado para verificar a aderência do modelo à proposta de enriquecimento semântico. A escala de pontuação proposta, conforme Quadro 11, foi utilizada para qualificar cada modelo de acordo com o grau de atendimento aos quesitos identificados no modelo proposto.

Em síntese, para a realização do estudo nos dois modelos de dados investigados foram usados os seguintes instrumentos:

1) Matriz TRESO versão 2.0, Quadro 12, contendo os critérios para detectar técnicas e os requisitos para a identificação de objetos em que foram gerados os registros da coleta dos dados realizada para cada modelo de dados;

2) A Escala de Pontuação, Quadro 11, usada para orientar a avaliação dos itens constantes na Matriz TRESO -

O resultado obtido em cada modelo foi comparado, de modo a identificar os pontos comuns e divergentes em cada um dos modelos de dados. Espera-se detectar a partir deste experimento, como a adoção de modelos de dados como o EDM e o BIBFRAME podem favorecer o enriquecimento semântico de objetos publicados em *Linked Data*, permitindo inclusive, a indicação de sugestões de melhoria para os modelos selecionados e para outros modelos.

### 5.2.1 Aplicação dos Critérios no BIBFRAME

O quadro 13 a seguir, exhibe na primeira coluna os critérios avaliados; na segunda coluna a pontuação atribuída ao BIBFRAME a cada um dos critérios, de acordo

com os parâmetros estabelecidos na Matriz TRESO; e a terceira coluna contém uma breve justificativa para a pontuação atribuída. Ressalta-se que foi detalhado na seção 5.3 as informações atinentes ao comportamento dos modelos em cada critério.

### Quadro 13 – Aplicação da Matriz ao BIBFRAME

(Continua)

| Critério de Avaliação  | Pontuação | Justificativa   |
|--|-----------|---|
| <p><b>1. Utilização de recursos de anotação semântica para o enriquecimento de objetos publicados em <i>Linked Data</i></b></p>                    | <p>2</p>  | <p>Enquanto na versão 1.0 do BIBFRAME o recurso de anotação estava relacionado a uma única classe específica, na versão 2.0 esta classe foi eliminada e os recursos de anotação foram inseridos como propriedades.</p> <p>O modelo de anotação, utilizando propriedades é considerado por Lira (2014) superior a anotação por <i>tags</i>, uma vez que o próprio atributo (propriedade) já explicita a relação existente entre a anotação e o recurso anotado.</p> <p>No entanto, a qualidade dessas anotações depende do conhecimento do usuário responsável, na medida em que não são utilizados vocabulários ou ontologias que possam facilitar o relacionamento entre os termos. Ainda segundo a autora, dificulta a reutilização da anotação e seu relacionamento com outros recursos na web.</p> <p>Deste modo, foi atribuída para este critério a <b>pontuação 02 – Atendimento Baixo.</b></p> |
| <p><b>2. Reuso de metadados para facilitar as atividades de publicação de dados em <i>Linked Data</i>, otimizando o trabalho do publicador</b></p> | <p>0</p>  | <p>Uma das propostas do BIBFRAME é ser capaz de acomodar um número abrangente de modelos de representação e padrões de metadados, para favorecer o compartilhamento e interligação dos relacionamentos no ambiente Web, sem limitar as atividades de catalogação.</p> <p>No entanto, ao invés de reutilizar propriedades já existentes em padrões de metadados, o BIBFRAME duplica propriedades do Dublin core para criar seu vocabulário a exemplo das propriedades <i>language</i>, <i>identifier</i> dentre outras.</p> <p>Mesmo que a reutilização de metadados seja descrita como uma premissa do modelo, a prática demonstra uma duplicação e não uma conexão das classes e propriedades com outros vocabulários, deste modo foi atribuída neste critério a <b>pontuação 0 - Inexistente</b></p>  |

Quadro 13 – Aplicação da Matriz ao BIBFRAME

(Continua)

| Critério de Avaliação   | Pontuação | Justificativa  |
|---|-----------|--|
| <p>3. Links entre as combinações semânticas dos dados e metadados com outros recursos da web.</p>   | <p>2</p>  | <p>Segundo o LOV (<i>Linked Open Vocabulary</i>) o BIBFRAME não é utilizado por nenhum <i>dataset</i> do LOD (<i>Linked Open Data</i>). As conexões com outros vocabulários são um link entrante com o Holding, que importou os dados do BIBFRAME, e dois links de saída, onde há uma relação dos metadados do BIBFRAME com o <b>DCterms</b> (termos do Dublin core) e o <b>RDF</b> (termos do vocabulário RDF).</p> <p>Deste modo foi atribuída a <b>pontuação – 2 atendimento baixo</b>.</p>   |
| <p>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.</p> | <p>3</p>  | <p>No BIBFRAME as relações são representadas por meio de grafos RDF, o que favorece uma maior liberdade e simplificação das representações.</p> <p>O vocabulário BIBFRAME é modelado a partir de um conjunto definido de elementos e atributos (classes e propriedades). Enquanto uma classe identifica um tipo de recurso BIBFRAME, as propriedades descrevem melhor um recurso e as relações entre os recursos são descritas em RDF/XML.</p> <p>Nesse critério foi atribuída a nota 3 – critério atendido, na medida em que os dados são modelados num formato semântico estruturado, como o RDF/XML, beneficiando sua manipulação por aplicações que consomem esses modelos de dados.</p> |
| <p>5. Utilização de ferramentas para o enriquecimento semântico.</p>  | <p>2</p>  | <p>O BIBFRAME fornece uma ferramenta para conversão de dados em formato MARC para o BIBFRAME. Futuramente serão fornecidas pela <i>Library of Congress</i> especificações e ferramentas e serviços de software para <i>download</i> para que as comunidades usuárias possam transformar e mover seus dados de MARC para BIBFRAME.</p> <p>As ferramentas atualmente disponibilizadas são fornecidas</p>   |

(Conclusão)

| Critério de Avaliação  | Pontuação | Justificativa   |
|--|-----------|---|
|  |           | <p>em Excel e servem para converter os dados de outras bibliotecas do MARC para BIBFRAME de forma semiautomática. Essas ferramentas não permitem a realização de outras atividades do processo de enriquecimento semântico, tais como: extração de dados, classificação de informações e anotação.</p> <p>Deste modo, foi atribuída a <b>nota 2 – Baixo atendimento</b>. O enriquecimento semântico é realizado de forma semiautomática por publicadores e usuários.</p>  |
| <b>6. Utilização de interface gráfica para apoiar a execução do processo de enriquecimento semântico e publicação de dados.</b>                      | <b>0</b>  | <p>A única ferramenta encontrada foi em aplicativo Excel para conversão de dados MARC para BIBFRAME. Não foi identificada nenhuma interface gráfica. Deste modo, foi atribuída a pontuação Nota <b>0 – critério inexistente</b>.</p>  |
| <b>7. Relações de sinonímia (equivalência), associação e hierarquia entre o metadado e o termo correspondente em outros vocabulários utilizados.</b> | <b>0</b>  | <p>No BIBFRAME as relações não são hierárquicas, associativas e de equivalência como ocorre em uma ontologia, e sim representadas por meio de grafos RDF. Apesar desses grafos poderem ser utilizados para representar diversos tipos de relações, no BIBFRAME a tipologia e natureza das relações existentes não estão explícitas.</p> <p>Apesar da utilização do RDF favorecer uma maior liberdade e simplificação das representações, não atende ao critério proposto por esta Matriz. Deste modo, foi atribuída a pontuação</p> <p><b>0 – critério inexistente.</b></p> |
| <b>Resultado</b>   | <b>9</b>  | <b>Atendimento Parcial</b>  |

Fonte: elaborado pela autora a partir da coleta de dados

### 5.2.2 Aplicação dos Critérios da Matriz ao EDM

O quadro 14 apresenta a aplicação dos critérios da Matriz TRESO no Modelo EDM.

Quadro 14 – Aplicação da Matriz no EDM

(Continua)

| Critério de Avaliação  | Pontuação | Justificativa   |
|--|-----------|---|
| Utilização de recursos de anotação semântica para o enriquecimento de objetos publicados em <i>Linked Data</i> | 3         | <p>O recurso de anotação adotado pelo EDM, utiliza tanto rotulagem simples por meio de tags, quanto anotação semântica e vinculação de objetos.</p> <p>O EDM adota o modelo de anotações baseado no padrão do modelo utilizado pela W3C, onde o objeto da anotação representa os recursos com o qual a anotação se refere.</p> <p>Uma das classes do EDM que normalmente recebe anotações é um objeto de patrimônio cultural fornecido (edm: ProvidedCHO) ou um recurso da web associado a ele, utilizando tags (literais) ou referências a vocabulários externos utilizando o RDF.</p> <p>Como modelo de dados EDM pode realizar a anotação semântica utilizando vocabulários controlados, de modo a permitir a conexão com dados de diferentes fontes foi atribuída a pontuação 3 – atendimento satisfatório.</p> |
| Reuso de metadados para facilitar as atividades de   | 4         | <p>O EDM possui suporte para recursos contextuais, incluindo conceitos de vocabulários controlados, permitindo a</p>  |

| Critério de Avaliação   | Pontuação | Justificativa   |
|---|-----------|---|
| <p>publicação de dados em <i>Linked Data</i>, otimizando o trabalho do publicador</p> |           | <p>reutilização de termos e relações.</p> <p>Em todas as classes do EDM existe o reaproveitamento de metadados de outros vocabulários, conforme abaixo:</p> <p>As propriedades da classe de objeto patrimônio cultural (Provided Cultural Heritage Object (edm:ProvidedCHO) reutilizam as propriedades do Dublin Core e OWL</p> <p>As propriedades da classe Recursos da web - Web Resource (edm:webResource) também reutilizam as propriedades da OWL</p> <p>As propriedades para a classe aggregation (ore:Aggregation) reutilizam propriedades do Dublin Core</p> <p>As propriedades para a classe Agent (edm:Agent) reutilizam propriedades do Dublin Core, SKOS, FOAF, RDA e OWL.</p> <p>As propriedades para a classe Place (edm:Place) reutilizam as propriedades da WGS, SKOS Dublin Core e OWL</p> <p>As propriedades para a classe <i>Timespan</i> ou <i>Period</i> (edm:TimeSpan) reutilizam as propriedades da SKOS, OWL e Dublin Core</p> <p>As propriedades para a classe Concept (skos:Concept) tem todas as propriedades do SKOS.</p> <p>Neste item foi atribuída a nota 4 – atendimento total, na medida em que o modelo de dados permite a reutilização de diferentes metadados em todas as propriedades do</p> |



| Critério de Avaliação  | Pontuação | Justificativa   |
|--|-----------|---|
|  |           | modelo.   |
| Links entre as combinações semânticas dos dados e metadados com outros recursos da web.  | 4         | <p>Além de reutilizar as propriedades de vários vocabulários e padrões, o EDM é utilizado em 10 <i>datasets</i> do LOD (<i>Linked Open Data</i>).</p> <p>Existem links de saída no LOV para padrões de metadados, vocabulários especializados e gerais, tais como FRBR, SKOS, RDFS, dentre outros.</p> <p>Como as conexões são realizadas por meio do RDF no âmbito do <i>Linked Data</i>, desta forma foi atribuída a nota 3 – atendimento satisfatório, porque o modelo prevê a integração entre as combinações semânticas dos dados e metadados com quaisquer <i>datasets</i> da web de dados.</p> |
| 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 | 4         | <p>O EDM é modelado por meio do uso de metadados transformados em URI, representados no padrão RDF/XML, o enriquecimento dos dados e a associação destes dados com os objetos culturais já existentes, são realizados para a conversão e a integração de dados armazenados em outras bases.</p> <p>Além disto o EDM utiliza conceitos da ontologia CIDOC CRM e suas generalizações com o propósito de realizar consultas altamente gerais para a um grande volume de dados.</p> <p>Para este critério foi atribuída a nota 4 – atendimento total,</p>   |

| Critério de Avaliação                                      | Pontuação | Justificativa  |
|--|-----------|--|
| Datasets do LOD.   |           | <p>na medida em que os dados são modelados num formato semântico estruturado, como OWL, bem como utiliza conceitos do CIDOC CRM o que beneficia sua manipulação por aplicações que consomem esses modelos de dados.</p>  |
| Utilização de ferramentas para o enriquecimento semântico. | 3         | <p>A Europeana desenvolveu uma pequena ferramenta de enriquecimento para fazer referências com URIs de vocabulário externos e buscar todos os dados multilíngues e semânticos anexados a um determinado conceito de vocabulário. Isso é facilitado quando o vocabulário alvo é representado pelo SKOS5, que também é o modelo EDM reutilizado para descrever dados conceituais.</p> <p>Da mesma forma, a Europeana realiza o enriquecimento automático com outros vocabulários e conjuntos de dados como GEMET6, GeoNames7 e DBpedia8 criando links para objetos na Europeana (Charles, Isaac e Freire (2014)).</p> <p>Nesse item foi atribuída a nota 3 – atendimento satisfatório já que o processo de enriquecimento semântico é feito de forma automática, utilizando ferramentas proprietárias para realizar anotações ou vinculação de dados e não por meio de software livre.</p> |

| Critério de Avaliação  | Pontuação | Justificativa   |
|--|-----------|---|
| Utilização de interface gráfica para apoiar a execução do processo de enriquecimento semântico e publicação de dados.                      | 0         | Não foi possível identificar a existência de interface gráfica para facilitar o processo de enriquecimento semântico e publicação de dados.<br><br>Deste modo foi atribuída a nota 0. Critério inexistente.   |
| Relações de sinonímia (equivalência), associação e hierarquia entre o metadado e o termo correspondente em outros vocabulários utilizados. | 4         | 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 tais como: DCterms:hasPart, DCterms:isPartOf, edm:hasType, edm:isSimilarTo, edm:type, owl:sameAs.<br><br>Neste item foi atribuída a nota 4 – atendimento total, na medida em que as propriedades utilizadas apresentam relações de sinonímia (equivalência), associação e hierarquia entre o metadado e o termo correspondente em outros vocabulários utilizados. |
| <b>RESULTADO</b>   | <b>22</b> | <b>ATENDIMENTO TOTAL</b>  |

Fonte: Elaborado pela autora a partir da coleta de dados

### 5.3 Resultados da Comparação dos Modelos de Dados

Esta seção descreve os resultados obtidos a partir do estudo comparativo da aderência dos modelos de dados EDM e BIBFRAME aos critérios, técnicas e recursos de

enriquecimento semântico propostos na Matriz TRESO. Os resultados desse estudo são apresentados a seguir.

### 5.3.1 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 os modelos utilizavam o recurso de anotação semântica, ou seja, como cada modelo adicionava 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. Contudo a forma adotada pelo EDM permite a utilização de links e vocabulários, atribuindo maior qualidade da anotação, se comparado ao BIBFRAME onde a anotação é realizada por meio de *tags* e propriedades que permitem o uso de texto livre e linguagem natural.

Conforme descrito na seção 2.1 os sistemas de anotação semântica devem utilizar sempre que possível, formatos padronizados, que permitem a integração entre recursos heterogêneos para que estes possam ser acessados e compartilhados simultaneamente. Neste sentido, segundo Uren et. Al (2005) a utilização de formatos padronizados é importante para o compartilhamento de anotações.

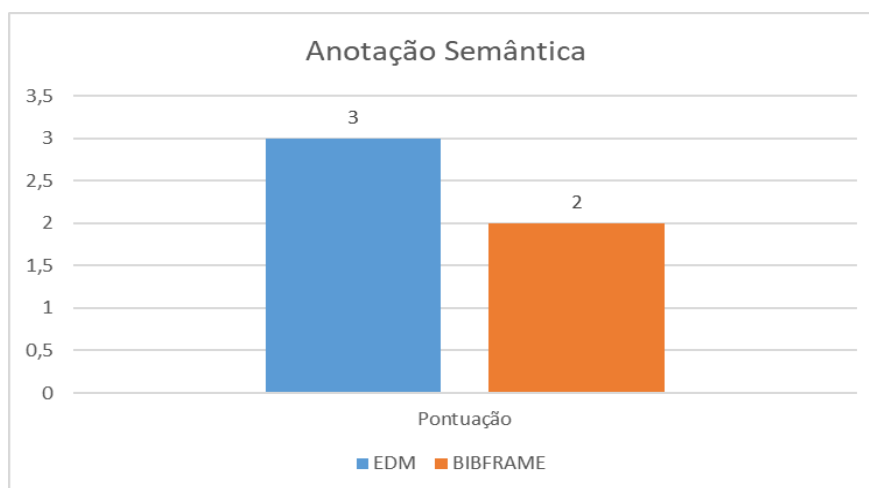
A utilização de linguagem natural no recurso de anotação, conforme adotado no BIBFRAME é criticada por Lira (2014), na medida em que pode promover equívocos, ambiguidade e dificuldade de interpretação por homens e máquinas, sendo considerada limitada se comparada à anotação semântica utilizando ontologias e vocabulários.

Já o EDM realiza a anotação por meio de *tags* (literais) ou 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.

Segundo Coneglian;Santarém (2017), o processo de anotação semântica é realizado pelo escritório da Europeana. As informações relevantes dos objetos culturais são relacionadas à tesouros e *datasets*, permitindo a ligação dos objetos de patrimônio cultural com as URIs existentes, de modo a conectar os dados com outras bases.

Como não foi possível observar se a forma de anotação adotada pelo EDM permite utilização de ontologias, que possibilitem aos sistemas a realização de inferências para localizar outros recursos, não foi atribuída a pontuação máxima neste item. O Gráfico 2, a seguir apresenta a pontuação obtida em cada modelo de dados no critério de anotação semântica.

**Gráfico 2 – Anotação semântica entre EDM e BIBFRAME**



Fonte: elaborado pela autora

Ressalta-se que a pontuação foi atribuída considerando a forma de atendimento de cada modelo ao critério de anotação semântica, em conformidade com os parâmetros estabelecidos na Matriz TRESO para esse critério. Em resumo, foi atribuída a nota dois para o BIBFRAME, porque as anotações foram realizadas em linguagem natural. Já para o EDM foi atribuída nota três, em virtude de este modelo não utilizar ontologias para realizar as anotações semânticas. Logo, há evidências de que apesar do EDM ser mais aderente neste critério do que o BIBFRAME, o mesmo não atingiu a excelência neste quesito por não adotar o recurso de anotação utilizando ontologias, considerado nesta pesquisa superior em relação aos demais tipos de anotação, conforme apontado na seção 2.1.3.1.1.

### 5.3.2 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 descrito anteriormente, apesar da proposta do BIBFRAME prever a acomodação de modelos e padrões de metadados, para favorecer o compartilhamento e interligação dos relacionamentos no ambiente Web, a prática demonstra que esse modelo cria propriedades idênticas ou similares às de outros padrões, ao invés de reutilizar essas propriedades, por meio da conexão entre os dados com o uso do RDF.

Na versão 2.0 do BIBFRAME apresentada no apêndice 1 deste documento, as propriedades *identifier*, *language*, *source* e *title*, possuem a mesma nomenclatura e função presentes no padrão Dublin Core, tratando-se de uma duplicação e não da reutilização da propriedade em seu vocabulário de origem.

Ainda no BIBFRAME 2.0 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 segundo Bizer (2009), dificulta a reutilização e, sobretudo a integração dos dados no *Linked Data*.

Enquanto no BIBFRAME há a duplicação de propriedades, a premissa do EDM é reutilizar termos e relações de diferentes vocabulários, por meio de referências, ou seja, o modelo utiliza as propriedades em seus vocabulários de origem. Deste modo, o EDM faz menção a elementos utilizados em outros esquemas como *Open Archives Object Reuse and Exchange Model* (OAI-ORE), Dublin Core e *Simple knowledge Organization System* – SKOS dentre outros.

Conforme pode ser observado na estrutura do EDM, apresentada no Apêndice 1 desta dissertação, em todas as classes do modelo é realizado o reaproveitamento de metadados de outros vocabulários. Observou-se que das 116 propriedades, 85 são reutilizadas de outros padrões, totalizando 73% das propriedades do modelo, conforme apontado no quadro 22 a seguir:

**Quadro 16 - Índice de Reutilização das propriedades do EDM**

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

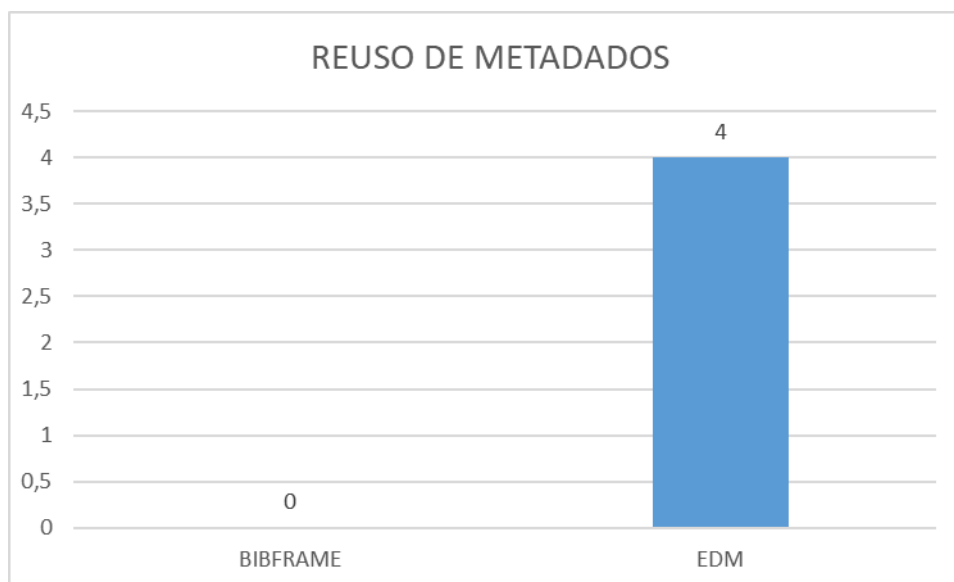
|           |  |               |  |
|-----------|--|---------------|--|
| (período) |  | 2 Dublin Core |  |
|           |  | 1 OWL         |  |

Fonte: Elaborado Pela autora a partir da coleta de dados

O diferencial entre estes dois modelos está principalmente na opção do EDM em reutilizar propriedades de vocabulários existentes, consolidados e disponíveis no LOD, ao invés de replicá-los. Essa opção permite maior facilidade para integração e consequente publicação de objetos uma vez que vários *datasets* utilizam alguns destes vocabulários.

O Gráfico 3 - Reutilização de Metadados EDM e BIBFRAME apresenta a pontuação obtida por cada modelo nesse critério. Como o EDM reutiliza metadados em todas as classes de modelo foi atribuída a nota quatro. Para o BIBFRAME foi atribuída a pontuação zero, tendo em vista que o mesmo realiza a duplicação e não a reutilização do dado em sua fonte de origem.

**Gráfico 3 - Reutilização de Metadados EDM e BIBFRAME**



Fonte: elaborado pela autora

Em relação a este quesito o estudo encontrou evidências que a forma de reutilização de metadados adotada pelo EDM, está alinhada aos princípios de publicação de



dados em *Linked Data*. Observou-se que isso ocorreu porque há a reutilização e integração de diferentes dados na web, permitindo a interoperabilidade entre os mesmos, o que justificou a pontuação máxima nesse critério.

### **5.3.3 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.

No BIBFRAME a responsabilidade por realizar as conexões entre os elementos do modelo BIBFRAME e os vocabulários de origem, como Dublin Core, FOAF, SKOS e outros vocabulários é dos *stakeholders* das bibliotecas que desejam converter seus dados para o formato BIBFRAME. Tal fato acaba por dificultar o processo de conexão entre os dados, já que se trata de uma decisão política de cada biblioteca e não uma premissa técnica do modelo.

De acordo com o LOV o BIBFRAME não é utilizado por nenhum *dataset* do LOD e suas conexões com outros vocabulários são um link entrante no qual o Holding importou os dados do BIBFRAME e dois links de saída onde há uma relação dos metadados do BIBFRAME com o DCterms (termos do Dublin Core) e o RDF (termos do vocabulário RDF).

Enquanto o BIBFRAME apresenta links apenas para os vocabulários do RDF e do Dublin Core, o EDM realiza a conexão com vários outros vocabulários disponíveis na Web Dados. Ainda segundo o LOV, o BIBFRAME apresenta dois links de saída para outros vocabulários. Já o EDM apresenta treze destes links de saída. O quadro 23 apresenta um comparativo entre conexões do BIBFRAME e EDM segundo o LOV.

**Quadro 17 – Comparativo entre as conexões do EDM e BIBFRAME no LOV**

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

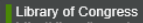
Fonte: elaborado pela autora


A FIG. 10 – Utilização do BIBFRAME no LOV e a FIG. 11 - Utilização do EDM no LOV apresentam as estatísticas de utilização de cada modelo no contexto da web semântica e do *Linked Open Data* publicados no site do LOV.

Figura 10 - Utilização do BIBFRAME no LOV

## BIBFRAME Vocabulary (bf)

### Metadata

|             |   |
|-------------|---|
| URI         | <a href="http://bibframe.org/vocab">http://bibframe.org/vocab</a>   |
| Namespace   | <a href="http://bibframe.org/vocab/">http://bibframe.org/vocab/</a>   |
| isDefinedBy | <a href="http://bibframe.org/vocab.rdf">http://bibframe.org/vocab.rdf</a>   |
| homepage    | <a href="http://www.loc.gov/bibframe">http://www.loc.gov/bibframe</a>   |
| Description | The BIBFRAME model and vocabulary consider resources that are cataloged as works with corresponding instances (physical and/or electronic). The metadata describing a work is an amalgamation of the some of the data that was formerly associated with the uniform title authority record combined with subject data that was associated with the bibliographic records. Thus each cataloging resource must have a work description, and, if the cataloging resource exists physically or electronically, one or more instance descriptions. @en |
| Language    |   |
| Creator     | <br><a href="http://dbpedia.org/resource/Library_of_Congress">http://dbpedia.org/resource/Library_of_Congress</a>  |
| Comment     | (2014-06-30) Ghislain Atemezing: There are still missing important metadata in this vocabulary, such as dct.description, dct.issued, and the creators.<br>(2015-04-03) Bernard Valant: Fixed namespace, vocabulary file URI and version history<br>(2016-09-22) Ghislain Atemezing: New version v2.0 in 2016  |



### Statistics

|            |   |
|------------|---|
| Classes    | 0 |
| Properties | 0 |
| Datatypes  | 0 |
| Instances  | 0 |

### Expressivity

[RDF](#) [RDFS](#)

### Tags

[Catalogs](#)

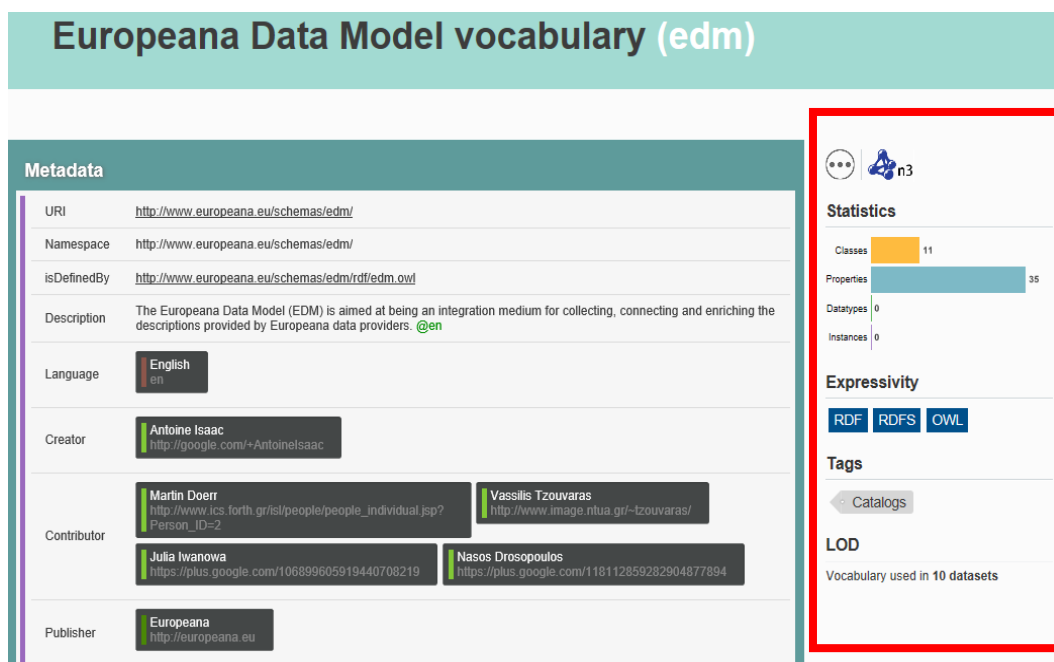
### LOD

Vocabulary used in 0 datasets

Fonte: Open Knowledge Foundation, 2018<sup>19</sup>

<sup>19</sup> Fonte: Open Knowledge Foundation. Disponível em <http://lov.okfn.org/dataset/lov/vocabs/bf>

Figura 11 - Utilização do EDM no LOV

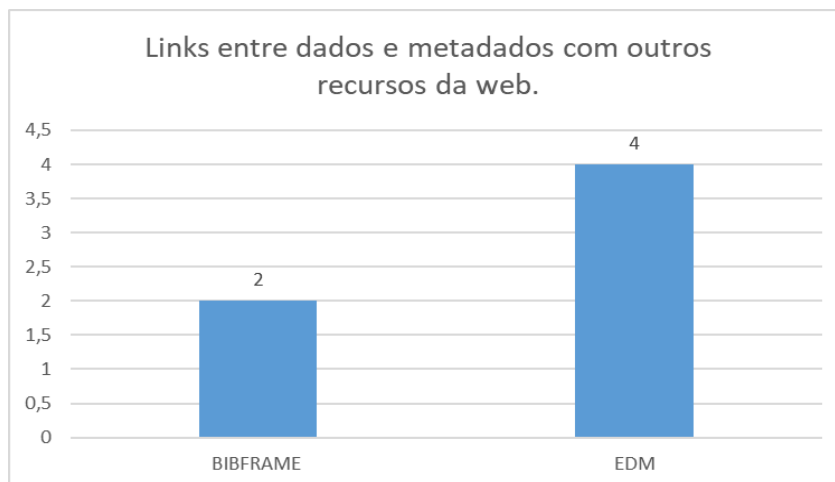


Fonte: Open Knowledge Foundation, 2018

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).

Conforme apresentado no **Gráfico 4 Link entre dados e metadados do EDM e BIBFRAME** com recursos da Web, ambos os modelos obtiveram pontuação nesse critério, uma vez que os dois modelos possuem links para outros *datasets* do LOD.

**Gráfico 4 – Link entre dados e metadados do EDM e BIBFRAME com recursos da Web**



Fonte: Elaborado pela autora

Em relação a este quesito, foi possível observar que o EDM atende totalmente ao critério estabelecido, por dois motivos. Primeiro porque o modelo realiza conexões com vários *datasets* do LOD. Segundo, porque ele 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 apenas com dois *datasets*.

#### **5.3.4 Critério 4 - Modelagem de dados num formato semântico estruturado**

A modelagem é uma importante ferramenta para a representação do conhecimento por que ajuda a reduzir os conflitos semânticos e a melhorar a documentação de sistemas. Furgeri (2006) e Silva (2011).

A partir da modelagem conceitual é possível identificar, extrair dados, compreender as relações entre conceitos e analisar as possibilidades de ligação e, ou separação semântica entre os conceitos de um determinado domínio, de modo a permitir sua interpretação por homens e máquinas. CAMPOS, 2004).

Através desse processo de modelagem é possível identificar as relações entre os objetos presentes em um mesmo contexto e eliminar os termos ambíguos que podem causar conflitos semânticos.

A etapa de modelagem é considerada uma boa prática para a publicação de dados abertos conectados, uma vez que define a forma de representação dos dados e os relacionamentos entre os mesmos, independentemente da aplicação e contexto em que os dados são utilizados.

No que se refere ao critério Modelagem de dados num formato semântico estruturado, tanto o EDM, quanto o BIBFRAME são modelados por meio do RDS/XML. 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 não são definidas a natureza e tipologia das relações, ou seja, não são definidas se as relações são do tipo todo-parte, gênero-espécie, etc.

O vocabulário BIBFRAME é modelado a partir de um conjunto definido de elementos e atributos (classes e propriedades). Enquanto a classe identifica um tipo de recurso, as propriedades descrevem este recurso e as relações entre os recursos são descritas em RDF/XML.

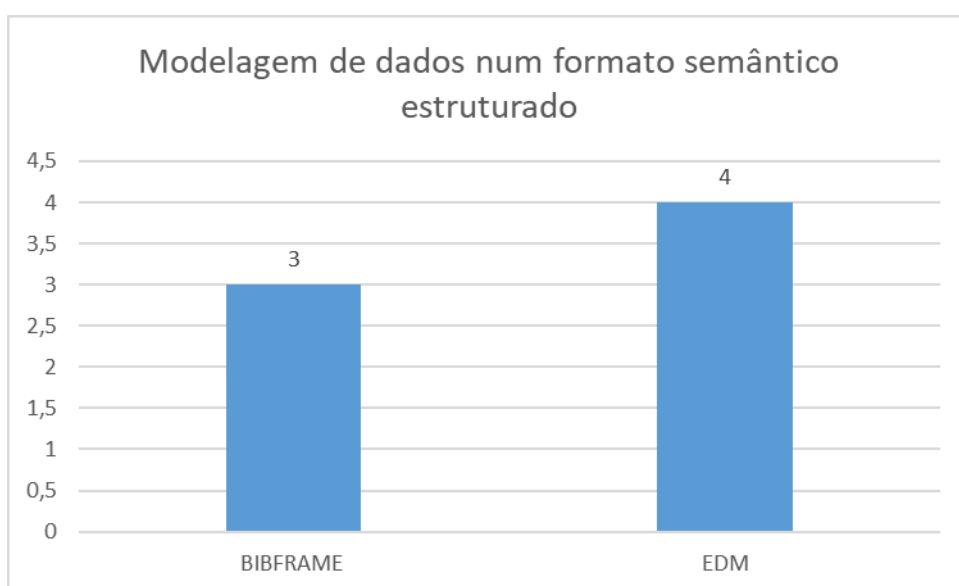
Já o EDM é modelado transformando seus metadados em URI, representados no padrão RDF/XML. O enriquecimento dos dados e a associação destes com os objetos culturais, são realizados para a conversão e a integração de dados de fontes existentes no LOD.

Além de utilizar o RDF/XML o EDM também utiliza a linguagem OWL e os conceitos da ontologia CIDOC CRM no processo de modelagem. 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 tem por objetivo tanto modelar os dados, quanto realizar consultas em grandes volumes de dados.

De acordo com Heckmann et. al., (2005), as ontologias possuem um importante papel no processo de modelagem e enriquecimento semântico. Segundo os autores as ontologias permitem a identificação dos tipos de relacionamento entre os conceitos, o que possibilita a realização de inferências sobre esses relacionamentos.

Conforme apresentado no **gráfico 5 - Modelagem de dados em um formato estruturado EDM e BIBFRAME**, nesse critério foi atribuída nota 4, para o EDM porque os dados são modelados num formato semântico estruturado, como OWL, bem como utiliza conceitos do CIDOC CRM, o que beneficia sua manipulação por aplicações que consomem esses modelos de dados. Já para o BIBFRAME foi atribuída nota três, uma vez que apesar dos dados serem modelados num formato semântico estruturado, como o RDF/XML, esse modelo não faz uso de ontologias para o desenvolvimento do seu processo de modelagem.

**Gráfico 5 – Modelagem de dados em um formato estruturado EDM e BIBFRAME**



Fonte: Elaborado pela autora

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 superior ao BIBFRAME, fato corroborado por Isotani; Bittencourt (2015). Segundo os autores as relações semânticas criadas pelas ontologias podem ser utilizadas pelo *Linked Data* para ampliar as conexões entre os recursos na web.

### **5.3.5 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. Segundo a Library of Congress <sup>20</sup>(2017), futuramente serão fornecidas especificações, ferramentas e serviços de *software* para que as comunidades usuárias possam fazer *download* e converter seus dados do formato MARC para BIBFRAME.

As ferramentas disponibilizadas, realizam o trabalho de modo semiautomático, permitem exclusivamente a conversão de MARC para BIBFRAME, não sendo possível realizar outras atividades de enriquecimento semântico, tais como: extração de dados, classificação, anotação, publicação e disseminação.

A Europeana 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 URIs de vocabulário externos, bem como são recuperados dados multilíngues e semânticos relacionados a um determinado conceito do vocabulário. Quando um vocabulário alvo é representado pelo SKOS, que também é reutilizado no EDM para descrever dados conceituais, todo o processo de enriquecimento é facilitado.

Da mesma forma, a Europeana realiza o enriquecimento automático com outros vocabulários e *datasets* como GEMET, GeoNames e DBpedia criando links para objetos na Europeana CHARLES, ISAAC, FREIRE (2014)

De acordo com Uren et al. (2005) a utilização de ferramentas automatizadas para a extração de conhecimento e anotação semântica é fundamental para reduzir o gargalo na aquisição de conhecimento, principalmente considerando as grandes coleções de documentos existentes na web.

Para a Bontcheva e Cunningham (2011), a importância da utilização de ferramentas tecnológicas é garantir a escalabilidade ou seja, a capacidade de processar milhões de documentos, bem como facilitar a personalização isto é, adaptar essas ferramentas a novos domínios e / ou ontologias. Contudo, 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,

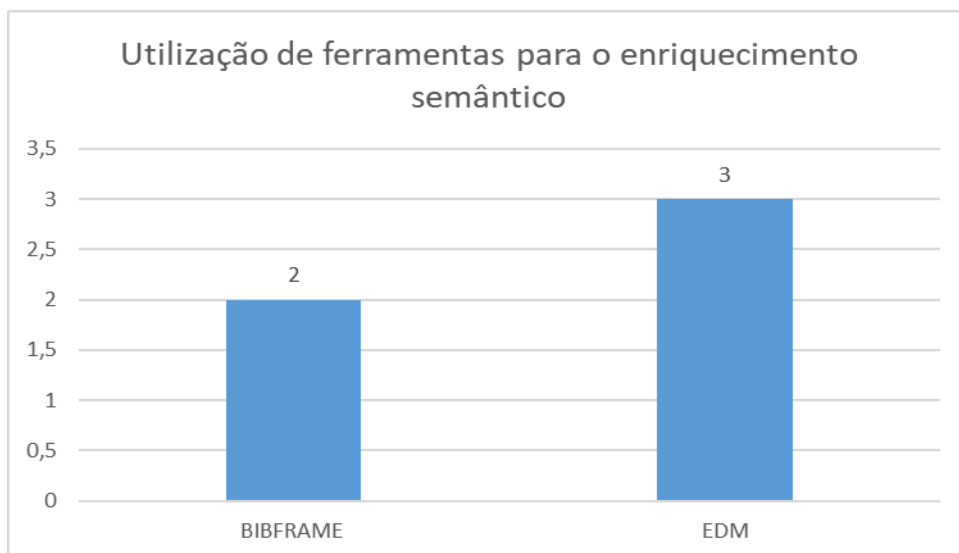
---

<sup>20</sup> <http://www.loc.gov/bibframe/>



conforme apresentado no gráfico 6, nesse item foi atribuída **nota 3 – atendimento satisfatório** - já que as anotações ou vinculação de dados são realizadas por meio de ferramentas proprietárias e não por meio de software livre.

**Gráfico 6 – Utilização de ferramentas para o Enriquecimento semântico**



Fonte: Elaborado pela autora

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.

### 5.3.6 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) para realizar um processo colaborativo é crucial fornecer aos publicadores, interfaces gráficas que simplifiquem o processo de anotação. Ainda segundo os autores a disponibilização de um ambiente de anotação dos documentos integrado com o ambiente de criação, leitura, compartilhamento e edição de informações poderia facilitar o processo de enriquecimento semântico. O *design* do sistema também

precisa facilitar a colaboração entre os usuários, permitindo que especialistas de diferentes campos possam contribuir e reutilizar documentos inteligentes.

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

### **5.3.7 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. Apesar de favorecer uma maior liberdade e simplificação das representações, não atende ao critério proposto por esta Matriz.

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 24, a seguir:

**Quadro 18 - Tipologia dos Relacionamentos do EDM**

| <b>Tipo de Relacionamento</b> | <b>Propriedade correspondente</b>                                |
|-------------------------------|--|
| Hierarquia                    | dcterms:hasPart,<br>dcterms:isPartOf,<br>edm:hasType<br>edm:type |

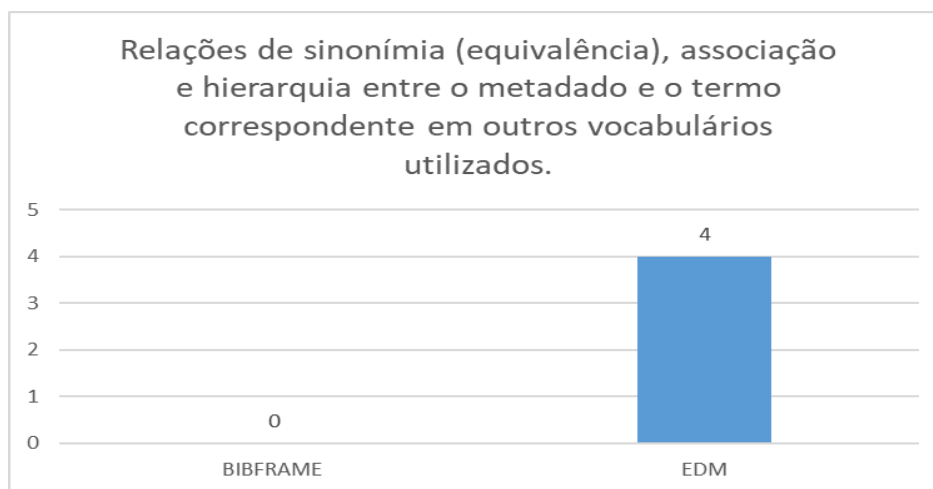
|                           |                 |
|---------------------------|-----------------|
| Associação                | edm:isSimilarTo |
| Sinonímia ou equivalência | owl:sameAs      |

Fonte: Elaborado pela autora

Neste item foi atribuído ao EDM nota quatro, 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 já que o mesmo é modelado em RDF e não explicita o tipo de relacionamento.

**Gráfico 7 – Relações entre o metadado e os termos de outros vocabulários**



Fonte: elaborado pela autora

Após a realização do estudo comparativo entre os modelos pode-se concluir, a partir das considerações apresentadas ao longo da análise de resultados, que o EDM apresenta grande parte dos requisitos e técnicas para o enriquecimento semântico, sendo mais aderente que o BIBFRAME aos aspectos avaliados pela Matriz TRESO.

Em relação a escala de pontuação, após a análise observou-se 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**.

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, sendo:

a)A utilização de vocabulários controlados e ontologias para facilitar a compreensão da anotação pelo usuário, quesito em que o BIBFRAME se mostrou inferior ao EDM, devido a adoção de linguagem natural, ao invés de vocabulários como o EDM;

b)A reutilização de metadados de outros modelos em seu contexto original. Critério em que novamente o EDM obteve pontuação superior ao BIBFRAME, em virtude do modelo fazer o reuso de propriedades de outros vocabulários em todas as suas classes;

c)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, principalmente se comparado ao BIBFRAME;

d) A modelagem de dados em um formato semântico estruturado, permitindo sua compreensão por homens e máquinas e sua consequente reutilização – ambos os modelos estão organizados de forma estruturada, contudo o uso de ontologias realizado pelo EDM pode ser considerado superior ao BIBFRAME, 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.

e) A explicitação dos relacionamentos entre os metadados do modelo e termos de outros vocabulários – Neste critério, como os tipos de relações entre os objetos e suas propriedades no BIBFRAME não são explicitados de modo, a permitir identificar se são hierárquicos, associativos ou equivalentes o mesmo não obteve pontuação neste item. Como o EDM possui algumas propriedades que indicam claramente o tipo de relação entre os objetos e suas propriedades o mesmo obteve pontuação total;

f) E a utilização de ferramentas para facilitar o trabalho do publicador e permitir a manipulação de grandes volumes de dados – Este critério se refere muito mais a quantidade de dados enriquecidos pelos modelos do que propriamente a qualidade dos dados no processo de publicação. Ambos os modelos possuem ferramentas, contudo as ferramentas utilizadas pelo EDM realizam o processo de forma mais automatizada e possuem mais funcionalidades do que a do BIBFRAME que só é utilizada para conversação de dados do MARC para o BIBFRAME de modo semiautomático.

Com tudo isso pode-se afirmar que o modelo EDM possui a maior parte dos elementos necessários para o enriquecimento semântico de objetos para sua publicação em *Linked Data*. Contudo, o mesmo ainda pode incorporar melhorias em seu modelo, através do 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 proposto por Lira(2014). Outro item que poderia trazer benefícios ao BIBFRAME é a utilização de ferramentas tecnológicas para facilitar o processo de enriquecimento semântico.

As discussões acerca do comportamento dos modelos em cada um dos critérios elencados na Matriz TRESO, indicou 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*.

## 6 CONSIDERAÇÕES FINAIS

A presente pesquisa buscou ampliar a discussão sobre o enriquecimento semântico de objetos para sua publicação em *Linked Data*. A questão que norteou a realização desta pesquisa foi: **como avaliar o enriquecimento semântico de objetos publicados na web através do *Linked Data*?**

Para responder ao problema de pesquisa, iniciou-se o estudo buscando suporte na literatura para compreender acerca dos conceitos de modelagem de dados, do enriquecimento semântico, e dos princípios para publicação de dados em *Linked Data*. Este levantamento bibliográfico foi realizado com o objetivo de construir um referencial teórico metodológico que possibilitasse o estudo dos objetos publicados em *Linked Data* do ponto de vista do enriquecimento semântico.

Entretanto, para atingir o objetivo geral e responder à pergunta desta pesquisa foi necessário desenvolver uma metodologia de análise que permitisse avaliar os modelos de dado, uma vez que durante a realização da pesquisa observou-se que não existia um instrumento de coleta de dados completo que permitisse verificar se de fato se os modelos de dados continham todos os elementos que pudessem favorecer o enriquecimento semântico. Deste modo, para atender ao objetivo pesquisa foi desenvolvida uma metodologia para a análise do enriquecimento semântico de objetos publicados na web através do *Linked Data*, por meio da Matriz TRESO.

Os objetivos específicos definidos no estudo foram atingidos, primeiramente porque a pesquisa foi baseada na análise e utilização de todos os conceitos, técnicas, recursos e ferramentas que de acordo com a literatura nacional e internacional da BCI e áreas correlatas favoreciam o processo de enriquecimento semântico para publicação de objetos em *Linked Data*.

Em segundo lugar, foi identificada a forma de publicação de dados adotada por dois modelos de dados o EDM e o BIBFRAME. Contudo, ressalta-se que esta análise foi baseada somente nos modelos selecionados, não só em virtude de sua representatividade e relevância no contexto do *Linked Data* e da credibilidade das instituições produtoras destes

modelos, mas também devido ao prazo para conclusão do estudo ser insuficiente para estender esta análise para outros modelos de dados.

Considerando, então, a análise a luz da matriz proposta, e da utilização de ambos os modelos no LOD, pode-se concluir que para um modelo de dados possa realizar o enriquecimento semântico de objetos para sua publicação em *Linked Data* é necessária a observância aos recursos e técnicas de enriquecimento semântico definidos na Matriz.

Acredita-se que os resultados obtidos foram satisfatórios e tornam viáveis a continuidade da pesquisa iniciada, na medida em que novos modelos podem ser desenvolvidos ou aprimorados com o presente trabalho. Deste modo, espera-se que a Matriz TRESO, resultante desta pesquisa 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* e enriquecidos semanticamente.

Ainda em relação às pesquisas futuras, pretende-se avançar no estudo dos modelos de dados e da proposição 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

AGUIAR, F. L. de; KOBASHI, N. Y. Organização e representação do conhecimento: perspectivas de interlocução interdisciplinar entre ciência da informação e arquivologia. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 14., 2013, Florianópolis. **Anais**. Florianópolis: UFSC, 2013. Disponível em: <<https://goo.gl/MfXBXA>>. Acesso em: 27 jul.2018.

ALVARENGA, L. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, v. 8, n. 15, 2003. Disponível em: < <https://goo.gl/g3LC4U> >. Acesso em: 27 jul.2018.

ARAKAKI, F. A. **Linked Data**: ligação de dados bibliográficos. 2016. 144 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016. Disponível em: <<https://goo.gl/tHdkcW>>. Acesso em: 27 jul.2018.

AUER, Sören et al. Introduction to linked data and its lifecycle on the web. In: RUDOLPH, SEBASTIAN et al. (Org.). **Reasoning web**: semantic technologies for intelligent data access. Lecture Notes in Computer Science. Berlin: Springer Berlin Heidelberg, 2013. p. 1– 90. Disponível em: < <https://goo.gl/kA5UV7>>. Acesso em: 03 ago. 2018.

AZEVEDO, P. C. N. **Uma proposta para visualização de Linked Data sobre enchentes na Bacia do Rio Doce**. 2014. 104 f. Dissertação (Mestrado em Sistemas de Informação e Gestão do Conhecimento) - Faculdade de Ciências Empresariais, Universidade FUMEC, Belo Horizonte, 2014.

BATISTA, M. G. R.; LÓSCIO, B. F. OpenSBBD: Usando Linked Data para Publicação de Dados Abertos sobre o SBBD. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS – SBBD, 2013, Recife. **Anais...** Recife: UFPE, 2013. Disponível em: <<https://goo.gl/LN4CQ5>>. Acesso em: 27 jul. 2018.

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://goo.gl/K37BJt>>\_. Acesso em: 27 jul. 2018.

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.

BONTCHEVA, K.; CUNNINGHAM, H. Semantic annotations and retrieval: manual, semiautomatic, and automatic generation. In: DOMINGUE, J.; FENSEL, D.; HENDLER, J. A. **Handbook of semantic web technologies**. Heidelberg: Springer-Verlag Berlin, 2011, p. 911-975.



CAMPOS, M. L. de A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**, Brasília, v. 33, n. 1, 2004. Disponível em: <https://goo.gl/XvD98e>. Acesso em 27 jul. 2018.

CHARLES, V. **Europeana and (many) linked open vocabularies**. 2012. 11 slides. Disponível em: <https://goo.gl/wwpojg>. Acesso em 27 jul. 2018.

CHARLES, V; Isaac, AHJCA ; Freire, N. linking libraries in the european library e europeana. In: IFLA SATELLITE MEETING LINKED DATA IN LIBRARIES, 2014, [s.l.]. **Anais...** Paris: IFLA, 2014.

CHEBOTKO, A. et al. An ontology-based multimedia annotator for the semantic web of language engineering. In: SHETH, A.; LYTRAS, M. (Ed.). **Semantic web-based information systems: state-of-the-art applications**. Hershey: Cybertech Publishing, 2007, p. 140-160.

CONEGLIAN, Caio Saraiva; SANTAREM, José Eduardo Segundo. Europeana no Linked Open Data: conceitos de webWeb 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, 2017. Disponível em: <<https://goo.gl/XTG56C>>. Acesso em 27 jul. 2018.

CUNHA, D.; LÓSCIO, B.; SOUZA, D. Linked Data: da web de Documentos para a web de Dados... 2011. In: ESCOLA REGIONAL DE COMPUTAÇÃO CEARÁ, MARANHÃO E PIAUÍ (ERCEMAPI), 5., 2011, Piauí. **Anais...** Piauí: [s.n.], 2011.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da Informação**, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978. Disponível em: <<https://goo.gl/RwTir6>>. Acesso em: 27 jul. 2018.

FURGERI, Sergio. **Representação de informação e conhecimento: estudo das diferentes abordagens entre a ciências da informação e a ciência da computação**. 2006. 161 f. Dissertação (Mestrado em Ciência da Informação) - Centro de Ciências Sociais Aplicadas, Universidade Estadual de Campinas, Campinas, 2006. Disponível em: <<https://goo.gl/5ja5GV>>. Acesso em: 27 jul. 2018.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 4. ed. São Paulo: Atlas, 1994. 207p.

GOMES, L. G. O.; PEREIRA, A. F. Identificação de pontos de controle no ciclo de desenvolvimento de produto por meio de modelagem conceitual e mapeamento da informação. **Gestão e Produção**. São Carlos, v. 21, n. 2, p. 257-269, 2014. Disponível em: <<https://goo.gl/JCczs7>>. Acesso em: 27 jul. 2018.

GRISOTO, A. P. **Um estudo acerca dos recursos audiovisuais no contexto do Linked data**. 2016. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2016.

GRISOTO, Ana Paula. **Um estudo acerca dos recursos audiovisuais no contexto do Linked Data**. 2016. 110 f. Dissertação (Mestrado) - Curso de Programa de Pós graduação em Ciência da Informação, Faculdade de Filosofia e Ciências - Unesp, Marília, 2016. Disponível em: < <https://goo.gl/tjDf4X> >. Acesso em: 03 ago. 2018.

GRUBER, T. Toward principles for the design of ontologies used for knowledge sharing. **International Journal Human-Computer Studies**. [s.l.], v. 43, p. 907-928, nov. 1995.

HALPIN, H.; LAVERENKO, V. Relevance feedback between hypertext and semantic search. In: SEMANTIC SEARCH WORKSHOP AT THE WORLD WIDE WEB CONFERENCE, 18., 2009. **Proceedings...** Madrid: [s.n.], 2009. Disponível em: <<https://goo.gl/sYU9nJ>>. Acesso em: 27 jul. 2018.

HASLHOFER, B.; ISAAC, A.. data. europeana. eu: the europeana linked open data pilot. In: **International Conference on Dublin Core and Metadata Applications**. 2011. p. 94-104.

HEATH, T. **Linked data-connect distributed data across the web**. 2012. Disponível em: <<http://linkeddata.org/>>.

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.

HODGE, G. **Systems of knowledge organization for digital libraries: beyond traditional authority files**. Washington, D.C.: The Digital Library Federation Council on Library Information Resources, 2000. Disponível em: <<https://goo.gl/4Ua9Xp>> Acesso em jul.2018.

HOHENSTEIN, U.; PLESSER, V. Semantic enrichment: a first step to provide database interoperability. In: WORKSHOP FÖDERIERTE DATENBANKEN, (1996). **Proceedings...** [s.l.: s.n.], (1996), p. 3-17.

HURFORD, James R.; HEASLEY, Brendan. **Curso de semântica**. Canoas: ULBRA, 1997.

HYLAND, B.; ATEMEZING, G.; VILLAZÓN-TERRAZAS, B. **Best Practices for Publishing Linked Data**. W3C Consortium. 2014 [Online]. Disponível em: < <https://goo.gl/VYdfDp/>>.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **Dados Abertos Conectados: Em busca da web do Conhecimento**. São Paulo: Novatec, 2015.

KENDALL, G. C.; FEIGENBAUM, L.; TORRES, E. **SPARQL Protocol for RDF**. 2008. Disponível em: <<https://goo.gl/4sJvGZ>>. Acesso em: 27 jul. 2018.

KLYNE, G.; CARROLL, J. J. **Resource description framework (RDF): concepts and abstract syntax**. W3C recommendation, W3C, Feb. 2004. Disponível em: < <https://goo.gl/dZUkoB>>. Acesso em: 03 ago. 2018.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de metodologia científica**. 3. ed. São Paulo: Atlas, 1991.

**Linked data platform use cases and requirements**. W3C Working Group Note 13 Marc. 2014. Disponível em: <<https://goo.gl/KT4GHP>>. Acesso em: 27 jul. 2018.

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://goo.gl/KKg6Pj>>. Acesso em: 27 jul. 2018.

LUHN, H. P. Key word-in-context index for technical literature (kwic index). **Journal of the Association for Information Science and Technology**, [s.l.], v. 11, n. 4, p. 288-295, 1960. Disponível em: <<https://goo.gl/vQLgb2>>. Acesso em 27 jul. 2018.

MARCONDES, C. H. Um modelo semântico de publicações eletrônicas. **Liinc em revista**, Rio de Janeiro, v. 7, n. 1, p. 82-103, 2011. Disponível em: <<https://goo.gl/WfG5B3>>. Acesso em 27 jul. 2018.

MARCONDES, C. H.; CAMPOS, M. L. de A. Ontologia e web semântica: o espaço da pesquisa em ciência da informação. **PontodeAcesso**, Salvador, v. 2, n. 1, p. 107-136, 2008. Disponível em: <<https://goo.gl/kgcqyo>>. Acesso em 27 jul. 2018.

MARCONDES, CH. "Linked Data" - dados interligados - e interoperabilidade entre arquivos, bibliotecas e museus na web. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 17, n. 34, p. 171-192, mai./ago. 2012. Disponível em: <<https://goo.gl/1qsLBC>>. Acesso em 27 jul. 2018.

MENDONÇA, R. R. de. **Uma abordagem para coleta e publicação de dados de proveniência no contexto de Linked Data**. 2013. 143 f. Tese (Mestrado em Informática) – Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013. Disponível em: <<https://goo.gl/PydDpD>>. Acesso em 27 jul. 2018.

MOREIRA, J. D. C. et al. Um sistema de enriquecimento semântico de perfil de usuário baseado em traços digitais para apoio à aprendizagem informal no contexto da saúde. **RENOTE**, Porto Alegre, v. 12, n. 1, jul. 2014. Disponível em: <<https://goo.gl/zFcAm2>>. Acesso em 27 jul. 2018.

MOURA, A. M. de C.; CAVALCANTE, Maria Claudia. Integração de recursos em portais semânticos e de documentos em organizações através da interoperabilidade de ontologias. In: SEMINÁRIO EM ONTOLOGIA NO BRASIL, 2. 2009. **Anais...** Rio de Janeiro: IME, 2009.

PEREIRA, Clayton Martins. **Uma abordagem para a publicação de dados abertos ligados obtidos a partir de bases de dados relacionais**. 2012. 157 f. Dissertação (Mestrado) - Curso de Engenharia Eletrônica e Computação, Instituto Tecnológico de Aeronáutica, São José dos Campos, 2012. Disponível em: <<https://goo.gl/JidX87>>. Acesso em: 03 ago. 2018.

PINHEIRO, J.C. **Processamento de consulta em um framework baseado em mediador para integração de dados no padrão de Linked Data**. 2011. 156 f. Tese (Doutorado) - Curso de Ciência da Computação, Universidade Federal do Ceará, Fortaleza, 2011. Disponível em: <<https://goo.gl/J3WnTx>>. Acesso em: 03 ago. 2018.

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

PRUD'HOMMEAUX, E.; SEABORNE, A. **Sparql query language for rdf**. 2008. Disponível em: < <https://goo.gl/yk3rTh> >. Acesso em: 27 jul. 2018.

PRUD'HOMMEAUX, Eric; SEABORNE, Andy. **SPARQL Query Language for RDF**. 2008. Disponível em: <<https://goo.gl/BVh2zq>>. Acesso em: 03 ago. 2018.

RAMALHO, R. A. S. **Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação**. 2006. 120 f. Dissertação (Mestrado em Ciência da Informação). Faculdade de Filosofia e Ciências – UNESP, Marília, 2006. Disponível em: < <https://goo.gl/FwGekN> >. Acesso em: 03 ago. 2018.

RAMALHO, R.A.S. bibframe: modelo de datos enlazados para bibliotecas. **Inf. Inf.**, Londrina, v. 21, n. 2, p. 292–306, maio/ago., 2016. Disponível em: <<https://goo.gl/Gkm8T1>> Acesso em jul. 2018.

ROCHA, A. L. C. **Um Modelo de Apresentação e Navegação de Linked Data para o Usuário Final**. 2014. 131 f. Dissertação (Mestrado em Ciência da Computação) - Universidade Estadual de Feira de Santana, Universidade Federal da Bahia, Salvador, 2014. Disponível em: <<https://goo.gl/gWKdgz>>. Acesso em: 27 jul. 2018.

ROCHA, R. P. FABRICO/CIÊNCIA: um ambiente Linked Data para o mapeamento da ciência. **Em Questão**, Porto Alegre, v. 18, p. 281-297, dez. 2012. Disponível em: <<https://goo.gl/JS6Yho>>. Acesso em: 27 jul. 2018.

SANTOS NETO, A. L. et al. Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. **Transinformação**, Campinas, v. 25, n. 1, p. 81-87, 2013. Disponível em: <<https://goo.gl/Mej6NH>>. Acesso em: 27 jul. 2018.

SAYÃO, L. F. Modelos teóricos em Ciência da Informação: abstração e método científico. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 82-91, jan./abr. 2001. Disponível em: < <https://goo.gl/9ncC5n>>. Acesso em: 27 jul. 2018.

SAYÃO, L.F.; MARCONDES, C.H. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**, Campinas, v. 20, n. 2, p. 133-148, 2008. Disponível em: < <https://goo.gl/mhc8M4> >. Acesso em: 27 jul. 2018.

SECRETARIA DE LOGÍSTICA E TECNOLOGIA DA INFORMAÇÃO – SLTI. MINISTÉRIO DO PLANEJAMENTO ORÇAMENTO E GESTÃO – MP. **CARTILHA técnica para publicação de dados abertos no Brasil v.1.0**. 2011. Disponível em: < <https://goo.gl/nQkdcY> >. Acesso em: 27 jul. 2018.

SHERA, J. H. Pattern, structure and conceptualization in classification. In: INTERNATIONAL STUDY CONFERENCE ON CLASSIFICATION FOR INFORMATION RETRIEVAL, 1957. **Proceedings...** London: ASLIB, 1957.

SILVA, A. R.; LIMA, G. A. B. O. A Teoria Clássica de Categorização e os Princípios. In: ENANCIB: políticas de informação para a sociedade, 11., 2011. Brasília. **Anais...** Brasília: UNB, 2011.

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: <<https://goo.gl/FJqtkC>>. Acesso em: 27 jul. 2018.

SILVA, J. L.M. da. **Uma abordagem para especificação de requisitos dirigida por modelos integrada ao controle da qualidade de aplicações Web**. 2011. 254 f. Tese (Doutorado em Engenharia de Sistemas e Computação) - COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011. Disponível em: <<https://goo.gl/RvZ8JW>>. Acesso em: 16 jun. 2017.

SILVA, Walison Dias da. **Anotação Semântica Automática do Currículo Lattes Utilizando Linked Open Data**. 2016. 132 f. Dissertação (Mestrado) - Curso de Sistemas de Informação e Gestão do Conhecimento, Faculdade de Ciências Empresariais, Fumec, Belo Horizonte, 2016. Disponível em: <<https://goo.gl/e1cGeG>>. Acesso em: 03 ago. 2018.

SOERGEL, Dt. **Knowledge Organization Systems: Overview**. 2014.

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

SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B. Towards a taxonomy of KOS: Dimensions for classifying Knowledge Organization Systems. **Knowledge organization**, [s.l.], v. 39, n. 3, p. 179-192, 2012. Disponível em: <<https://goo.gl/3rZtvX>>. Acesso em: 27 jul. 2018.

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.

THE EDM MAPPING GUIDELINES. **Europeana Data Model – Mapping Guidelines v2.2**. 2014. Disponível em: <<https://goo.gl/WbsPjt>>. Acesso em: 16 mar. 2017.

TOMITA, M. **Tecnologias para aplicação da Web Semântica nas unidades de informação. Londrina**, 2015. Dissertação (Mestrado em Ciência da Informação) - Centro de Educação, Comunicação e Artes, Universidade Estadual de Londrina, Londrina, 2015. Disponível em: <<https://goo.gl/vXPDU2>>. Acesso em: 27 jul. 2018.

TOSIN, Thyago de Melo. **Um modelo para integração de informações de bases de dados abertos, com uso de ontologias**. 2016. 107 f. Dissertação (Mestrado) - Curso de <https://goo.gl/mjt92a>, Escola Politécnica, Universidade do Vale do Rio dos Sinos, São Leopoldo, 2016. Disponível em: <<https://goo.gl/mJT92a>>. Acesso em: 03 ago. 2018.

TRISTÃO, A. M. D.; FACHIN, G. R. B.; ALARCON, O. E. Sistemas de classificação facetada e tesouros: instrumentos para organização do conhecimento. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 161-171, maio/ago. 2004. Disponível em: <<https://goo.gl/5tJv1K>>. Acesso em: 27 jul. 2018.

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.

Vandenbussche, P.Y., Ateazing, G.A., Poveda-Villalón, M., Vatant, B. Linked open vocabularies (lov): a gateway to reusable semantic vocabularies on the web. **Semantic Web (Preprint)**, [s.l.], p. 1–16, (2015). Disponível em: < <https://goo.gl/cxPSXs>>. Acesso em 27 jul. 2018.

WINER, D.; ROCHA, I. E. Europeana: um projeto de digitalização e democratização do patrimônio cultural europeu. **Patrimônio e Memória**, 2013. Disponível em: < <https://goo.gl/4CtPPi>>. Acesso em 03 ago. 2018.

## APÊNDICE

### AÊNDICE 1 – PROPRIEDADES DO EDM: PROVIDEDCHO

Quadro 19 - Propriedades do Objeto Patrimônio Cultural (Provided Cultural Heritage Object (edm:ProvidedCHO))

| PROPRIEDADE                  | DESCRIÇÃO   |
|------------------------------|---|
| <b><i>dc:contributor</i></b> | Propriedade originada do Dublin Core, utilizada para descrever contribuidores para o CHO. Sempre que possível, vincular o identificador do contribuidor a partir de uma fonte/catálogo de autoridade.   |
| <b><i>dc:coverage</i></b>    | Originada do Dublin Core, utilizada para designar tópicos espaciais ou temporais da classe de objeto cultural CHO.  |
| <b><i>dc:creator</i></b>     | Também oriunda do Dublin Core, utilizada para designar o criador do CHO. Sempre que possível deve-se fornecer o identificador do criador a partir de uma fonte de autoridade. Esta propriedade deve ser repetida, caso existam múltiplos criadores. |
| <b><i>dc:date</i></b>        | Utilizado para uma data significativa do CHO. Podem ser utilizada com as sub-propriedades <code>dcterms:created</code> or <code>dcterms:issued</code> , ambas oriundas do Dublin Core.  |
| <b><i>dc:description</i></b> | Descrição do objeto de patrimonio cultural CHO  |
| <b><i>dc:format</i></b>      | Utilizado para os termos geralmente aplicados para indicar o formato do objeto do patrimônio cultural ou o formato de arquivo de um objeto nato digital.  |
| <b><i>dc:identifier</i></b>  | Identificador do objeto de patrimônio cultural original tal como ISBN, ISSN, URI etc.   |
| <b><i>dc:language</i></b>    | O idioma do objeto de patrimônio cultural de texto e também para outros tipos de objeto se houver um aspecto de linguagem. Obrigatório para objetos textuais, fortemente recomendado para outros tipos de objeto com algum elemento de idioma.      |
| <b><i>dc:publisher</i></b>   | O nome da editora do objeto. Se possível, fornecer o identificador do editor a partir de uma fonte de autoridade.   |
| <b><i>dc:relation</i></b>    | O nome ou identificador de um recurso relacionado, geralmente usado para outros objetos relacionados.   |
| <b><i>dc:rights</i></b>      | Utilizado para denominar o detentor de direitos do CHO, ou para informações de direitos mais gerais.  |

| PROPRIEDADE                          | DESCRIÇÃO   |
|--------------------------------------|---|
| <b><i>dc:source</i></b>              | Recurso relacionado no qual o recurso descrito seja derivado total ou parcialmente, da fonte do CHO original.   |
| <b><i>dc:subject</i></b>             | Assunto do objeto de patrimônio cultural.   |
| <b><i>dc:title</i></b>               | O título do CHO. As traduções exatas do título podem ser fornecidas usando os atributos de idioma apropriados   |
| <b><i>dc:type</i></b>                | A natureza ou gênero do CHO. Os termos devem ser retirados de um vocabulário controlado.  |
| <b><i>dcterms:alternative</i></b>    | Título alternativo do CHO, incluindo abreviaturas ou traduções mesmo que não sejam exatas.  |
| <b><i>dcterms:conformsTo</i></b>     | Padrão estabelecido cujo CHO está em conformidade.  |
| <b><i>dcterms:created</i></b>        | Data de criação do objeto de patrimônio cultural.   |
| <b><i>dcterms:extent</i></b>         | Tamanho ou duração do CHO.  |
| <b><i>dcterms:hasFormat</i></b>      | Recurso relacionado ao CHO que é substancialmente o mesmo que o CHO, mas com formato diferente.   |
| <b><i>dcterms:hasPart</i></b>        | Recurso incluído física ou logicamente no CHO.  |
| <b><i>dcterms:hasVersion</i></b>     | Recurso posterior que é uma versão, edição ou adaptação do CHO demonstrando mudanças substanciais no conteúdo e não no formato.   |
| <b><i>dcterms:isFormatOf</i></b>     | Outro recurso que é substancialmente o mesmo que o CHO, mas em outro formato.   |
| <b><i>dcterms:isPartOf</i></b>       | Recurso no qual o CHO está incluído fisicamente ou logicamente. Esta propriedade pode ser usada para objetos que fazem parte de uma hierarquia e será usada para suportar uma exibição apropriada no portal. Para esse efeito, será necessário fornecer uma referência como valor. Parte de |
| <b><i>dcterms:isReferencedBy</i></b> | Outro recurso que faz referência, cita ou aponta para o CHO.  |
| <b><i>dcterms:isReplacedBy</i></b>   | Outro recurso que suplanta, desloca ou substitui o CHO.   |
| <b><i>dcterms:isRequiredBy</i></b>   | Outro recurso relacionado que requer o CHO para suportar sua função, entrega ou coerência.  |



| PROPRIEDADE                           | DESCRIÇÃO   |
|---------------------------------------|---|
| <b><i>dcterms:issued</i></b>          | Data de emissão formal ou publicação do CHO.  |
| <b><i>dcterms:isVersionOf</i></b>     | Outro, recurso anterior do qual o CHO é uma versão, edição ou adaptação, demonstrando mudanças substanciais no conteúdo e não no formato.   |
| <b><i>dcterms:medium</i></b>          | O material ou veículo físico do CHO.  |
| <b><i>dcterms:provenance</i></b>      | Uma declaração de mudanças na propriedade e custódia do CHO desde a sua criação. Significativo para autenticidade, integridade e interpretação.   |
| <b><i>dcterms:references</i></b>      | Outros recursos referenciados, citados ou de outra forma apontados pelo CHO.  |
| <b><i>dcterms:replaces</i></b>        | Um recurso relacionado que é substituído, deslocado ou substituído pelo CHO.  |
| <b><i>dcterms:requires</i></b>        | Outro recurso requerido pelo recurso descrito para suportar sua função, entrega ou coerência.   |
| <b><i>dcterms:spatial</i></b>         | Características espaciais do CHO. Isto é, o que o CHO representa ou representa em termos de espaço.   |
| <b><i>dcterms:tableOfContents</i></b> | Lista de sub-unidades do CHO.   |
| <b><i>dcterms:temporal</i></b>        | Características temporais do CHO.   |
| <b><i>edm:currentLocation</i></b>     | A localização geográfica cujos limites atualmente incluem o CHO. Se o nome de um repositório, construção, site ou outra entidade for usado, ele deve incluir uma indicação de sua localização geográfica.       |
| <b><i>edm:hasMet</i></b>              | O identificador de um agente, um lugar, um período de tempo ou qualquer outra entidade identificável que o CHO pode ter "encontrado" em sua vida.   |
| <b><i>edm:hasType</i></b>             | O identificador de um conceito, ou uma palavra ou frase de um vocabulário controlado (thesaurus etc), identificando o tipo de CHO   |
| <b><i>edm:incorporates</i></b>        | O identificador de outro recurso que está incorporado no CHO descrito.  |
| <b><i>edm:isDerivativeOf</i></b>      | O identificador de outro recurso a partir do qual o CHO descrito foi derivado.  |
| <b><i>edm:isNextInSequence</i></b>    | O identificador do objeto anterior em que ambos os objetos fazem parte do mesmo recurso geral. Use isso para objetos que fazem parte de uma hierarquia ou seqüência para garantir a exibição correta no portal. |
| <b><i>edm:isRelatedTo</i></b>         | O identificador ou o nome de um conceito ou outro recurso ao qual o CHO descrito está relacionado.  |

| PROPRIEDADE                      | DESCRIÇÃO   |
|----------------------------------|---|
| <b><i>isRepresentationOf</i></b> | O identificador de outro objeto do qual o CHO descrito é uma representação.   |
| <b><i>edm:isSimilarTo</i></b>    | O identificador de outro recurso ao qual o CHO descrito é semelhante.   |
| <b><i>edm:isSuccessorOf</i></b>  | O identificador de um recurso ao qual o CHO descrito é um sucessor.   |
| <b><i>edm:realizes</i></b>       | Se o CHO descrito for de tipo edm: PhysicalThing pode realizar um objeto de informação.   |
| <b><i>edm:type</i></b>           | O valor deve ser um dos tipos aceitos pela Europeia, pois suportará a funcionalidade do portal.   |
| <b><i>owl:sameAs</i></b>         | Use para apontar para a sua própria (representação de conteúdo vinculado) a representação do objeto, caso já tenha sido digitado um identificador URI para ele. |

Fonte: Adaptado pela autora de *The EDM mapping guidelines*, (2014).

#### Quadro 20 - Propriedades da classe Recursos da web - Web Resource (edm:webResource)

| PROPRIEDADE               | DESCRIÇÃO  |
|---------------------------|--|
| <b>dc:creator</b>         | Utilizado para o criador do recurso da web. Se possível, fornecer o identificador do criador a partir de uma fonte de autoridade.  |
| <b>dc:description</b>     | Utilizado para uma conta ou descrição desta representação digital  |
| <b>dc:format</b>          | Utilizado para o formato desta representação digital.  |
| <b>dc:rights</b>          | Utilizado para o nome do detentor dos direitos desta representação digital, se possível, ou para informações de direitos mais gerais.  |
| <b>dc:source</b>          | Um recurso relacionado a partir do qual o recurso da web é derivado, no todo ou em parte.  |
| <b>dcterms:conformsTo</b> | Um padrão estabelecido para o qual o recurso da web está em conformidade.  |
| <b>dcterms:created</b>    | Data de criação do recurso da web.   |
| <b>dcterms:extent</b>     | O tamanho ou a duração do recurso digital.   |
| <b>dcterms:hasPart</b>    | Um recurso que está incluído física ou logicamente no recurso da web.  |
| <b>dcterms:isFormatOf</b> | Outro recurso que é substancialmente o mesmo que o recurso da web, mas em outro formato.   |
| <b>dcterms:isPartOf</b>   | Um recurso no qual o recurso da web está incluído fisicamente ou logicamente. Esta propriedade pode ser usada para recursos da web que fazem parte de uma hierarquia. As hierarquias podem ser |

| PROPRIEDADE                 | DESCRIÇÃO  |
|-----------------------------|--|
|                             | representadas como hierarquias de Organismos fornecidos ou hierarquias de recursos da web, mas não ambas ao mesmo tempo.   |
| <b>dcterms:issued</b>       | Data da emissão formal ou publicação do recurso da web.  |
| <b>edm:isNextInSequence</b> | Onde um CHO possui vários recursos da web, mostrado por múltiplas instâncias da propriedade edm: hasView no ore: Agregation esta propriedade pode ser usada para mostrar a seqüência dos objetos. Cada recurso da web (além do primeiro na seqüência) deve usar essa propriedade para fornecer a URI do recurso anterior na seqüência.   |
| <b>edm:rights</b>           | O valor neste elemento irá indicar os direitos de autor, uso e direitos de acesso que se aplicam a essa representação digital. É altamente recomendável que um valor seja fornecido para esta propriedade para cada instância de um recurso da web. A declaração de direitos especificada ao nível do recurso da web irá "substituir" a declaração especificada no nível da agregação. |
| <b>owl:sameAs</b>           | Forneça o URI de outra representação web do mesmo recurso.   |

Fonte: Adaptado pela autora de *The EDM mapping guidelines*, (2014).

**Quadro 21 - Propriedades para a classe aggregation (ore:Aggregation)**

| PROPRIEDADE              | DESCRIÇÃO  |
|--------------------------|--|
| <b>edm:aggregatedCHO</b> | O identificador do objeto de origem. Pode ser um URI de dados aberto vinculado completo ou um identificador interno.   |
| <b>edm:dataProvider</b>  | Nome ou identificador do provedor de dados do objeto (ou seja, a organização que fornece dados para um agregador). Os identificadores não estarão disponíveis até que a Europeia tenha implementado seu perfil de organização. |
| <b>edm:hasView</b>       | O URL de um recurso web que é uma representação digital do CHO. Este pode ser o próprio objeto de origem no caso de um objeto de patrimônio cultural gerado em formato digital.  |
| <b>edm:isShownAt</b>     | O URL de uma visão web do objeto no contexto completo da informação.   |
| <b>edm:isShownBy</b>     | O URL de uma visão web do objeto.  |
| <b>edm:object</b>        | O URL de uma representação do CHO que será usado para gerar visualizações para uso no portal Europeia.   |

| PROPRIEDADE         | DESCRIÇÃO  |
|---------------------|--|
| <b>edm:provider</b> | O nome ou o identificador do provedor do objeto (ou seja, a organização que fornece dados diretamente para o Europeana).   |
| <b>dc:rights</b>    | Esta propriedade deve ser aplicada ao edm: webResource ou o edm: ProvidedCHO. É incluído aqui para a conversão de dados da ESE, onde não se sabe a que objeto se aplicam os direitos.                          |
| <b>edm:rights</b>   | Esta é uma propriedade obrigatória e o valor dado aqui deve ser a declaração de direitos que se aplica à representação digital como dada, quando esses recursos não são fornecidos com seus próprios direitos. |
| <b>edm:ugc</b>      | Esta é uma propriedade obrigatória para objetos gerados ou criados pelo usuário que foram coletados pela fonte do público ou atividade de projeto.   |

Fonte: Adaptado pela autora de *The EDM mapping guidelines*, (2014).

**Quadro 22 - Propriedades para a classe Agent (edm:Agent)**

| PROPRIEDADE                           | DESCRIÇÃO  |
|---------------------------------------|--|
| <b>skos:prefLabel</b>                 | A forma preferida do nome do agente. Embora o número máximo de ocorrências esteja definido em 1, ele pode ser interpretado como 1 por tag de idioma. |
| <b>skos:altLabel</b>                  | Formas alternativas do nome do agente.   |
| <b>skos:note</b>                      | Uma nota sobre o agente.   |
| <b>dc:date</b>                        | Uma data significativa associada ao agente.  |
| <b>dc:identifier</b>                  | Um identificador do agente.  |
| <b>dcterms:hasPart</b>                | Referência a um agente que faz parte do agente que está sendo descrito (por exemplo, uma parte de uma corporação).                                   |
| <b>dcterms:isPartOf</b>               | Referência a um agente do qual o agente descrito faz parte.  |
| <b>edm:begin*</b>                     | A data em que um agente nasceu ou foi criado.  |
| <b>edm:end*</b>                       | Data em que um agente faleceu ou foi extinto.  |
| <b>edm:hasMet</b>                     | Referência a outra entidade que o agente "conheceu" em um sentido amplo. Por exemplo, uma referência a uma classe de lugar.                          |
| <b>edm:isRelatedTo</b>                | Referência a outras entidades, particularmente outros agentes, com quem o agente está relacionado em um sentido genérico.                            |
| <b>foaf:name</b>                      | O nome do agente como uma seqüência textual simples.   |
| <b>rdaGr2:biographicalInformation</b> | Informações relativas à vida ou história do agente.  |
| <b>rdaGr2:dateOfBirth</b>             | A data em que o agente (pessoa) nasceu.  |
| <b>rdaGr2:dateOfDeath</b>             | A data em que o agente (pessoa) morreu.  |
| <b>rdaGr2:dateOfEstablishment</b>     | A data em que o agente (órgão corporativo) foi estabelecido ou fundado.  |
| <b>rdaGr2:dateOfTermination</b>       | A data em que o agente (órgão corporativo) foi encerrado ou dissolvido.  |
| <b>rdaGr2:gender</b>                  | O gênero com o qual o agente identifica.   |
| <b>rdaGr2:placeOfBirth</b>            | Local (cidade, província, estado e / ou país) em que uma pessoa nasceu.  |
| <b>rdaGr2:placeOfDeath</b>            | Local (cidade, província, estado e / ou país) em que uma pessoa faleceu.   |
| <b>rdaGr2:professionOrOccupation</b>  | Profissão ou ocupação em que o agente trabalha ou trabalhou.   |
| <b>owl:sameAs</b>                     | Outro URI do mesmo agente.   |

Fonte: Adaptado pela autora de The EDM mapping guidelines, (2014).

**Quadro 23 - Propriedades para a classe Place (edm:Place)**

| PROPRIIDADE          | DESCRIÇÃO   |
|----------------------|---|
| wgs84_pos:lat        | A latitude de um objeto espacial (graus decimais).  |
| wgs84_pos:long       | A longitude de um objeto espacial (graus decimais).   |
| wgs84_pos:alt        | A altitude de um objeto espacial (metros decimais acima da referência).   |
| skos:prefLabel       | A forma preferencial do nome do lugar. O máximo é 1 por tags de idioma.   |
| skos:altLabel        | Formas alternativas do nome do local.   |
| skos:note            | Informações relativas ao local.   |
| dcterms:hasPart      | Referência a um lugar que faz parte do local que está sendo descrito.   |
| dcterms:isPartOf     | Referência a um lugar onde o lugar descrito é parte do outro.   |
| edm:isNextInSequence | Pode ser usado para representar uma sequência de entidades e locais ao longo do tempo, e. Utilizada para objetos que fazem parte de uma hierarquia ou sequência para garantir a exibição correta no portal. |
| owl:sameAs           | URI de um lugar.  |

Fonte: Adaptado pela autora de *The EDM mapping guidelines*, (2014).

**Quadro 24 - Propriedades para a classe Timespan ou Period (edm:TimeSpan)**

| PROPRIIDADE      | DESCRIÇÃO   |
|------------------|---|
| skos:prefLabel   | O formulário preferido do nome do período ou período. Embora o número máximo de ocorrências esteja definido em 1, ele pode ser interpretado como 1 por tag de idioma. |
| skos:altLabel    | Formas alternativas do nome do período ou intervalo de tempo.   |
| skos:note        | Informações relativas ao intervalo de tempo ou período.   |
| dcterms:hasPart  | Referência para um intervalo de tempo que faz parte do período de tempo descrito.   |
| dcterms:isPartOf | Referência a um intervalo de tempo em que o período de tempo descrito é uma parte.  |
| edm:begin        | A data em que o período de tempo começou.   |
| edm:end          | A data em que o período de tempo terminou.  |

|                             |   |
|-----------------------------|---|
| <b>edm:isNextInSequence</b> | Pode ser usado para representar uma sequência de períodos de tempo. Utilizado para objetos que fazem parte de uma hierarquia ou sequência para garantir a exibição correta no portal. |
| <b>owl:sameAs</b>           | O URI de um período de tempo  |

Fonte: Adaptado pela autora de *The EDM mapping guidelines*, (2014).

**Quadro 25 - Propriedades para a classe Concept (skos:Concept)**

| PROPRIIDADE   | DESCRIÇÃO   |
|---|---|
| <b>skos:prefLabel</b>   | A forma preferida do nome do conceito. A instância máxima é 1 por tag de idioma.  |
| <b>skos:altLabel</b>  | Formas alternativas do nome do conceito.  |
| <b>skos:broader</b>   | O identificador de um conceito mais amplo no mesmo tesouro ou vocabulário controlado.   |
| <b>skos:narrower</b>  | O identificador de um conceito mais estreito.   |
| <b>skos:related</b>   | O identificador de um conceito relacionado  |
| <b>skos:broadMatch</b><br><b>skos:narrowMatch</b><br><b>skos:relatedMatch</b> | O identificador de conceitos de correspondência mais amplos, mais estreitos ou relacionados de outros esquemas conceituais.                     |
| <b>skos:exactMatch</b><br><b>skos:closeMatch</b>                              | O identificador de conceitos próximos ou exatamente correspondentes de outros esquemas conceituais.   |
| <b>skos:note</b>  | Informações relativas ao conceito.  |
| <b>skos:notation</b>  | A notação em que o conceito é representado. Isso pode não ser palavras em linguagem natural para alguns sistemas de organização do conhecimento |
| <b>skos:inScheme</b>  | O URI de um esquema conceitual  |

Fonte: Adaptado pela autora de *The EDM mapping guidelines*, (2014).

**Quadro 26 - Propriedades do BIBFRAME**

| CLASSE                  | DESCRIÇÃO   |
|-------------------------|---|
| <b>AdminMetadata:</b>   | Metadados do metadado, especificamente informação da proveniência do recurso  |
| <b>Agent:</b>           | Entidade que tem um papel no recurso como uma Pessoa ou Organização   |
| <b>AppliedMaterial:</b> | Substância física ou química aplicada a um material de base de um recurso   |
| <b>Arrangement:</b>     | Informações sobre a organização e arranjo de uma coleção de itens. Por exemplo, para arquivos de computador, organização e informação arranjo pode ser a estrutura de arquivos e ordenar sequência de um arquivo; para materiais visuais, esta informação pode ser como uma coleção é organizado. |
| <b>AspectRatio:</b>     | Relação proporcional entre a largura de uma imagem e sua altura.  |
| <b>BaseMaterial:</b>    | Material físico subjacente de um recurso  |



| CLASSE                           | DESCRIÇÃO   |
|----------------------------------|---|
| <b>BookFormat:</b>               | Resultado de dobrar uma folha impressa para formar uma reunião de folhas.   |
| <b>Capture:</b>                  | Informação sobre lugar e data associado com a captura (isto é, a gravação, filmar, etc.) do conteúdo de um recurso.   |
| <b>Carrier:</b>                  | Categorização refletir a forma do meio de armazenamento e alojamento de um transportador.   |
| <b>Cartographic:</b>             | Conteúdo que representa a totalidade ou parte da terra, qualquer corpo celeste, ou lugar imaginário em qualquer escala.                                     |
| <b>Classification:</b>           | Sistema de codificação e organização de materiais de acordo com o seu assunto.  |
| <b>Collection:</b>               | Agregação de recursos, geralmente reunidos artificialmente.   |
| <b>ColorContent:</b>             | Características de cor de um recurso, por exemplo, preto e branco, multicolor, etc.   |
| <b>Content:</b>                  | Categorização refletindo a forma fundamental de comunicação em que o conteúdo é expresso e o sentido humano através do qual ele se destina a ser percebida. |
| <b>Contribution:</b>             | Agente e papel no que respeita ao recurso que está sendo descrito.  |
| <b>CopyrightRegistration:.</b>   | Direitos autorais ou informações de registro Depósito Legal   |
| <b>CoverArt:</b>                 | Ilustração da capa de um recurso.   |
| <b>DigitalCharacteristic:</b>    | Especificação técnica relativa à codificação digital de texto, imagem, áudio, vídeo e outros tipos de dados em um recurso.                                  |
| <b>Dissertation:</b>             | Informações sobre um trabalho apresentado como parte dos requisitos formais de um grau acadêmico.   |
| <b>Emulsion:</b>                 | Suspensão de produtos químicos sensíveis à luz usado como um revestimento sobre um microfilme ou microfichas, por exemplo, halogeneto de prata.             |
| <b>EnumerationAndChronology:</b> | Numeração ou outra enumeração e datas associadas problemas ou elementos detidos.  |
| <b>Event:</b>                    | Algo que acontece em um determinado momento e local, como uma performance, fala ou evento esportivo, que é documentada por um recurso.                      |
| <b>Extent:</b>                   | Número e tipo de unidades e / ou subunidades que compõem um recurso.  |
| <b>FontSize:</b>                 | Tamanho do tipo dos utilizados para representar os caracteres e símbolos em um recurso.   |
| <b>Frequency:</b>                | Informações sobre intervalos em que as partes de um recurso produzidos em série ou as atualizações para um recurso de integração são emitidos.              |

| CLASSE                   | DESCRIÇÃO   |
|--------------------------|---|
| <b>Generation:</b>       | Relações entre um transportador original e o portador de uma reprodução feita a partir do original.   |
| <b>GenreForm:</b>        | Indicação do programa ou processo utilizado para gerar a descrição de aplicação por uma transformação em particular.  |
| <b>Identifier:</b>       | Token ou o nome que está associado com um recurso, tal como uma URI ou um ISBN.   |
| <b>Illustration:</b>     | Informações sobre o conteúdo destinado a ilustrar um recurso.   |
| <b>Instance:</b>         | Resource refletindo uma encarnação individual, material de uma obra.  |
| <b>IntendedAudience:</b> | A informação que identifica o público alvo ou alvo específico ou nível intelectual para o qual o conteúdo descrito é considerado apropriado. Também é usado para gravar os níveis de interesse e motivação e as características especiais do aluno. |
| <b>Issuance:</b>         | Informações sobre se um recurso é emitido em uma ou mais partes, a forma como ele é atualizado, e seu término previsto.   |
| <b>Item:</b>             | único exemplo de uma instância.   |
| <b>Language:.</b>        | Entidade idioma   |
| <b>Layout:</b>           | Arranjo de texto, imagens, notação tátil, etc., em um recurso.  |
| <b>Media:</b>            | Categorização refletindo o tipo geral do dispositivo de intermediação necessário para ver, brincar, correr, etc., o conteúdo de um recurso.   |
| <b>Mount:</b>            | material físico utilizado para o suporte ou apoio para o qual foi ligado o material de base de um recurso.  |
| <b>MusicEnsemble:</b>    | Conjunto para o qual uma obra musical é apropriado.   |
| <b>MusicFormat:</b>      | Disposição para o conteúdo de um recurso que é apresentada sob a forma de notação musical.  |
| <b>MusicInstrument:</b>  | Instrumento para que uma obra musical é apropriada.   |
| <b>MusicMedium:</b>      | Declaração Resumo do meio de um trabalho musical.   |
| <b>MusicVoice:</b>       | Voz para a qual uma obra musical é apropriada   |
| <b>Notation:</b>         | Informações sobre o alfabeto, roteiro, ou sistema símbolo usado para transmitir o conteúdo do recurso, incluindo roteiros especializados, tipos, notação tátil, notação movimento, e notação musical.   |
| <b>Note:</b>             | Informações, geralmente em forma de texto, em atributos de um recurso ou algum aspecto de um recurso.   |
| <b>Place:</b>            | Localização geográfica  |
| <b>Polarity:</b>         | Relação das cores e tons em uma imagem com as cores e tons do objeto reproduzido.   |

| CLASSE                           | DESCRIÇÃO  |
|----------------------------------|--|
| <b>ProductionMethod:</b>         | Processo utilizado para produzir um recurso  |
| <b>ProjectionCharacteristic:</b> | Especificação técnica referente à projeção de um filme de cinema.  |
| <b>ProvisionActivity:</b>        | Informações sobre o agente ou local relativa à publicação, impressão, distribuição, emissão, liberação, ou a produção de um recurso. |
| <b>ReductionRatio:</b>           | Tamanho de uma micro-imagem em relação ao original a partir do qual foi produzido.   |
| <b>Review:</b>                   | Avaliação de um recurso.   |
| <b>SoundCharacteristic:</b>      | Especificação técnica referente à codificação de som, de um recurso.   |
| <b>SoundContent:</b>             | Indicação do facto de a produção de som é uma parte integrante do recurso.   |
| <b>Source:</b>                   | Recurso do qual valor ou o rótulo veio ou foi derivado.  |
| <b>Status:</b>                   | Designação da validade ou a posição de alguma coisa, por exemplo, se algo está incorreto ou disponível.                              |
| <b>Sublocation:</b>              | lugar específico dentro da entidade exploração em que o item está localizado ou disponibilizado.                                     |
| <b>Summary:</b>                  | Descrição do conteúdo de um recurso, como um resumo, resumo, etc.  |
| <b>SupplementaryContent:</b>     | Índice, bibliografia, apêndice, etc. destina-se a complementar o conteúdo principal de um recurso                                    |
| <b>TableOfContents:</b>          | Tabela de conteúdo de um recurso.  |
| <b>Temporal:</b>                 | cronológica.   |
| <b>Title:</b>                    | A informação do título, relativa a um recurso: título mail, título traduzido, ou variante do título.                                 |
| <b>Topic:</b>                    | Conceito ou área de conhecimento.  |
| <b>Unit:</b>                     | As unidades em que um valor é expresso.  |
| <b>UsageAndAccessPolicy:</b>     | Mapa geral de subsídios e restrições de acesso a um recurso, incluindo a retenção, reprodução, acesso e concessão de empréstimos.    |
| <b>VideoCharacteristic:</b>      | especificação técnica relacionada com a codificação de imagens de vídeo em um recurso.   |
| <b>Work:</b>                     | Resource refletindo uma essência conceitual de um recurso de catalogação.  |

Fonte: ARAKAKI 2016 Baseado em *Library of Congress* (201-), p.85