

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE BIBLIOTECOLOGIA E COMUNICAÇÃO  
DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO

Maína Guimarães Rymsha

***Linked Open Data* como ambiente para a publicação de dados abertos  
bibliográficos e da ciência na web**

Porto Alegre

2013

Maína Guimarães Rymsza

***Linked Open Data* como ambiente para a publicação de dados abertos  
bibliográficos e da ciência na web**

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção de grau de Bacharel em Biblioteconomia, pela Faculdade de Biblioteconomia e Comunicação, da Universidade Federal do Rio Grande do Sul.

Orientador: Dr. Rafael Port da Rocha

Porto Alegre

2013

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**

Reitor: Carlos Alexandre Netto

Vice-Reitor: Rui Vicente Oppermann

**FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO**

Diretora: Ana Maria Mielniczuk de Moura

Vice-Diretor: André Iribure Rodrigues

**DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO**

Chefe: Maria do Rocio Fontoura Teixeira

Chefe-Substituto: Valdir Jose Morigi

**COMISSÃO DE GRADUAÇÃO DO CURSO DE BIBLIOTECONOMIA**

Coordenadora: Samile Andréa de Souza Vanz

Vice-Coordenadora: Glória Isabel Sattamini Ferreira

**Dados Internacionais de Catalogação na Publicação – CIP**

R995l Rymsza, Maína Guimarães  
*Linked Open Data* como ambiente para a publicação de dados abertos bibliográficos e da ciência na web / Maína Guimarães Rymsza – 2013.  
66 f.

Monografia (Trabalho de Conclusão de Curso) –  
Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Curso de Biblioteconomia, 2013.

Orientador: Prof. Dr. Rafael Port da Rocha

1. *Linked Open Data*. 2. Dados abertos. 3. Web. 4. Web Semântica. I. Rocha, Rafael Port da. II. Título.

**Departamento de Ciências da Informação**

Rua Ramiro Barcelos, 2705

Bairro Santana - Porto alegre – RS

CEP 90035-007

Telefone: (051) 3308-5097

E-mail: fabico@ufrgs.br

MAÍNA GUIMARÃES RYMSZA

***Linked Open Data* como ambiente para a publicação de dados abertos  
bibliográficos e da ciência na web**

Trabalho de Conclusão de Curso  
apresentado como requisito parcial para a  
obtenção de grau de Bacharel em  
Biblioteconomia, pela Faculdade de  
Biblioteconomia e Comunicação, da  
Universidade Federal do Rio Grande do  
Sul.

Aprovado em: \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_.

**BANCA EXAMINADORA**

---

Prof. Dr. Rafael Port da Rocha (Orientador)  
Universidade Federal do Rio Grande do Sul

---

Profa. Dra. Ana Maria Mielniczuk de Moura  
Universidade Federal do Rio Grande do Sul

---

Prof. Dr. Rodrigo Silva Caxias de Sousa  
Universidade Federal do Rio Grande do Sul

Aos meus pais,  
por todo o amor, incentivo e paciência.

## **AGRADECIMENTOS**

À Universidade Federal do Rio Grande do Sul, por possibilitar um ensino gratuito e de qualidade.

Ao meu orientador, professor Rafael, por contribuir imensamente para este trabalho. Agradeço pela ajuda, incentivo e pelo aprendizado proporcionado nas orientações para a elaboração deste trabalho.

Aos amigos queridos da Biblioteca do Tribunal de Justiça do Rio Grande do Sul (TJ-RS): Eliana, Adri, Magda, Ziza, Vera, Karina, Aline e Lidiane, pelas ótimas tardes, cafés e conversas.

Aos amigos queridos da Biblioteca do Ministério Público do Trabalho do Rio Grande do Sul (MPT-RS): Carmen, Sachi, Sônia, Vera e Maurício, pelos aprendizados e pelas conversas.

Aos amigos da FABICO, que tornaram o período da faculdade tão agradável. Agradeço em especial às colegas: Juliani, Dinara, Karina, Indiara, Silvana, Carmen, Márcia e Bruna.

A todos meus amigos, que sempre estiveram presentes na minha vida, em especial: Jaime, Ana, Dani, Luiza e Márcia.

A toda minha família, por sempre acreditarem em mim e por me incentivarem.

Aos meus irmãos Vinícius, Gabriel e Camila pelo carinho, apoio e companheirismo.

E finalmente, agradeço especialmente aos meus pais, Mario e Janete, por todo amor, dedicação, incentivo e paciência nos períodos mais difíceis da faculdade.

Obrigada a todos!

*“A evolução das profissões da informação está ligada, muito de perto,  
ao progresso da ciência e da tecnologia da informação.”*

**Le Coadic**

## RESUMO

Apresenta estudo sobre os dados abertos ligados, *Linked Open Data*, como ambiente para a publicação de dados abertos bibliográficos e dados abertos da ciência na web. Utiliza como metodologia para realização da pesquisa a abordagem qualitativa e de cunho exploratório, com a finalidade de se conhecer mais a respeito do tema. Desenvolve o estudo com base na revisão de literatura sobre o tema e análise dos principais projetos desenvolvidos na publicação de dados bibliográficos e dados da ciência utilizando *Linked Open Data*. No referencial teórico, apresenta a conceitualização e contextualização dos Dados Abertos, Web Semântica, *Linked Open Data* e suas aplicações em Bibliotecas e na Ciência. No desenvolvimento da pesquisa, analisa aspectos pré-determinados de cinco projetos de dados abertos ligados, sendo três sobre publicação de dados abertos bibliográficos: The British National Bibliography Linked Data, Virtual International Authority File e Europeana Linked Data, e dois projetos sobre dados abertos da ciência: Linked Open Data University of Muenster e Eagle-I. Conclui que *Linked Open Data* é um ambiente interessante e promissor para a publicação de dados bibliográficos e de dados da ciência na web, trazendo os benefícios de sua utilização para as instituições, para os profissionais da Ciência da Informação e para a sociedade.

**Palavras-chave:** *Linked Open Data*. Web. Dados abertos bibliográficos. Dados abertos da ciência.



## **ABSTRACT**

Presents study on Linked Open Data as an environment for publishing open data open bibliographic data and science on the web. Used as a methodology for research, the qualitative approach and exploratory, with the purpose of knowing more about the subject. Develops study based on literature review on the subject and analysis of main projects in publishing bibliographic data and science data using Linked Open Data. In the theoretical framework, presents the conceptualization and contextualization of Open Data, Semantic Web and Linked Open Data in Libraries and their applications in science. The development of research, analyzes aspects predetermined five projects linked open data, three on publishing open bibliographic data: The British National Bibliography Linked Data, Virtual International Authority File and Europeana Linked Data, and two projects on the open data science: Linked Open Data University of Muenster and Eagle-I. Concludes that Linked Open Data is an interesting and promising environment for publishing bibliographic data and science data on the web, bringing the benefits of its use for institutions, for professionals in Information Science and society.

**Keywords:** Linked Open Data. Web. Open bibliographic data. Open data science.

## LISTA DE ILUSTRAÇÕES

<b>Figura 1</b> –	URI, URL e URN	29
<b>Figura 2</b> –	Exemplo de Estrutura em Triplas	30
<b>Figura 3</b> –	Diagrama de Nuvem dos Dados Abertos Ligados em 2007	36
<b>Figura 4</b> –	Diagrama de Nuvem dos Dados Abertos Ligados em 2009	36
<b>Figura 5</b> –	Diagrama de Nuvem dos Dados Abertos Ligados em 2011	37
<b>Figura 6</b> –	Esquema 5 estrelas para Open Data	38

## LISTA DE QUADROS

<b>Quadro 1</b>	–	Projetos Seleccionados	19
<b>Quadro 2</b>	–	Objetivos e Coleta de Dados	21
<b>Quadro 3</b>	–	The British National Bibliography (BNB) Linked Data	43
<b>Quadro 4</b>	–	Virtual International Authority File (VIAF)	47
<b>Quadro 5</b>	–	Europeana Linked Data	48
<b>Quadro 6</b>	–	Linked Open Data University of Münster (LODUM)	52
<b>Quadro 7</b>	–	Eagle-I	54
<b>Quadro 8</b>	–	Comparativo das Ontologias utilizadas nos Projetos	55
<b>Quadro 9</b>	–	Comparativo das Bases de Dados ligadas aos Projetos	57

## LISTA DE SIGLAS

<b>BIBO</b>	Bibliographic Ontology
<b>BNB</b>	The British National Bibliography
<b>Brapci</b>	Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação
<b>CDD</b>	Classificação Decimal de Dewey
<b>CRM</b>	Conceptual Reference Model
<b>ERO</b>	Eagle-I Research Resource Ontology
<b>FOAF</b>	Friend Of a Friend
<b>HTTP</b>	Hypertext Transfer Protocol
<b>ISBD</b>	International Standard Bibliographic Description
<b>LCSH</b>	Library of Congress Subject Headings
<b>LD</b>	Linked Data
<b>LISA</b>	Library and Information Science Abstracts
<b>LOD</b>	Linked Open Data
<b>LODUM</b>	Linked Open Data University of Muenster
<b>OAI-ORE</b>	Open Archives Initiative Object Reuse and Exchange
<b>OBI</b>	Ontology for Biomedical Investigation
<b>OCLC</b>	Online Computer Library Center
<b>OCRe</b>	Ontology for Clinical Research
<b>OF</b>	Open Format
<b>OL</b>	Open License
<b>OWL</b>	Web Ontology Language
<b>PRISM</b>	Publishing Requirements for Industry Standard Metadata
<b>RDA</b>	Resource Description and Access
<b>RDF</b>	Resource Description Framework
<b>RE</b>	Machine-Readable
<b>SciELO</b>	Scientific Electronic Library Online
<b>SKOS</b>	Simple Knowledge Organization System
<b>SO</b>	Sequence Ontology
<b>SWO</b>	Software Ontology
<b>TEACH</b>	Teaching Core Vocabulary

<b>Uberon</b>	Uber-anatomy Ontology
<b>UFRGS</b>	Universidade Federal do Rio Grande do Sul
<b>URI</b>	Uniform Resource Identifier
<b>URL</b>	Uniform Resource Locator
<b>URN</b>	Uniform Resource Name
<b>VIAF</b>	Virtual International Authority File
<b>W3C</b>	Consórcio World Wide Web

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
1.1	PROBLEMA DE PESQUISA	16
1.2	JUSTIFICATIVA	16
1.3	OBJETIVOS	17
<b>1.3.1</b>	<b>Objetivo Geral</b>	<b>17</b>
<b>1.3.2</b>	<b>Objetivos Específicos</b>	<b>17</b>
<b>2</b>	<b>METODOLOGIA</b>	<b>19</b>
2.1	ABORDAGEM E TIPO DE PESQUISA	19
2.2	CORPUS DA PESQUISA	19
2.3	INSTRUMENTOS DE COLETA DE DADOS	20
2.4	PROCEDIMENTOS DE COLETA DE DADOS	22
<b>3</b>	<b>DADOS ABERTOS</b>	<b>23</b>
<b>4</b>	<b>WEB SEMÂNTICA</b>	<b>27</b>
4.1	UNIFORM RESOURCE IDENTIFIER (URI)	28
4.2	RESOURCE DESCRIPTION FRAMEWORK (RDF)	30
4.3	WEB ONTOLOGY LANGUAGE (OWL)	31
4.4	ONTOLOGIAS PARA WEB SEMÂNTICA	32
<b>4.4.1</b>	<b>Friend of a Friend (FOAF)</b>	<b>33</b>
<b>4.4.2</b>	<b>GeoNames</b>	<b>34</b>
<b>5</b>	<b>LINKED OPEN DATA (LOD)</b>	<b>35</b>
5.1	BIBLIOTECAS LIGANDO DADOS	40
<b>5.1.1</b>	<b>Projetos de Dados Bibliográficos em Linked Open Data</b>	<b>41</b>
5.1.1.1	The British National Bibliography (BNB) Linked Data	41
5.1.1.2	Virtual International Authority File (VIAF)	44
5.1.1.3	Europeana Linked Data	47
5.2	A CIÊNCIA LIGANDO DADOS	49
<b>5.2.1</b>	<b>Projetos de Dados da Ciência em Linked Open Data</b>	<b>50</b>
5.2.1.1	Linked Open Data University of Münster (LODUM)	50
5.2.1.2	Eagle-I	52
5.3	ONTOLOGIAS E BASES DE DADOS	55
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>58</b>
	<b>REFERENCIAS BIBLIOGRÁFICAS</b>	<b>61</b>

<b>APÊNDICE A – Fichamento dos Dados</b>	<b>65</b>
<b>APÊNDICE B – Formulário de Coleta de Dados</b>	<b>66</b>

## 1 INTRODUÇÃO

Vivemos em uma sociedade em rede, em que grande parte da população está conectada à Internet. A facilidade para a utilização das ferramentas da Internet e sua grande adesão mundial popularizou sua utilização, formando uma sociedade de indivíduos conectados, que produzem e utilizam as informações disponibilizadas na grande massa de dados que forma a web. O crescimento da web acontece de forma exponencial. Diariamente, grandes volumes de dados são publicados sem nenhum tipo de tratamento dessas informações, o que, possivelmente, ocasionará dificuldades na recuperação da informação.

Entre as iniciativas mais recentes na web, está a dos chamados Dados Abertos. Dados Abertos é um movimento pela publicação de dados na Internet com acesso livre, ou seja, de dados que podem ser visualizados e utilizados livremente por qualquer pessoa que tenha conexão com a Internet. A disponibilização dos dados abertos para a sociedade tem como principal finalidade a disseminação de informação que possa ser relevante para o usuário. Como exemplo de dados abertos, temos os dados abertos governamentais, que são as informações relativas aos órgãos do governo publicadas de forma a garantir a transparência das informações públicas, os dados abertos bibliográficos, que se referem às informações extraídas de registros bibliográficos, acervos de bibliotecas e catálogos de autoridades, e os dados abertos da ciência, que correspondem à publicação das informações provenientes de estudos e pesquisas científicas, como dos dados coletados em uma pesquisa, para que estejam disponíveis para reutilização, propiciando a evolução na ciência, através de estudos colaborativos e descobertas científicas.

A fim de possibilitar a recuperação dos dados abertos disponíveis na web é essencial a utilização de ferramentas durante o tratamento e a organização das informações publicadas. Nesse contexto, surge o *Linked Open Data* como uma iniciativa para a publicação de dados abertos na web com a intenção de contribuir para a organização e recuperação da informação. *Linked Open Data* são recomendações de melhores práticas para a publicação de dados abertos na web e que visam promover uma maior interoperabilidade e usabilidade dos dados armazenados, possibilitando que o acesso e a utilização dos dados ocorram de forma mais eficiente. A web foi concebida para ligar documentos, já *Linked Open*



*Data* traz para esta web a interligação dos dados e para tal utiliza como infraestrutura a Web Semântica.

O presente estudo analisa *Linked Open Data* como mecanismo para a publicação dos dados abertos na web, através do desenvolvimento de uma pesquisa bibliográfica que tem por objetivo trazer uma revisão de literatura sobre os principais conceitos e abordagens que já foram publicados sobre o tema em questão e apresentar os principais projetos que foram desenvolvidos utilizando *Linked Open Data*, com enfoque nos dados abertos bibliográficos e dados abertos da ciência.

### 1.1 PROBLEMA DE PESQUISA

O que é *Linked Open Data* e de que forma as instituições estão utilizando este ambiente para a publicação dos dados abertos bibliográficos e dados abertos da ciência na web?

### 1.2 JUSTIFICATIVA

Em primeiro lugar, a pesquisa justifica-se devido o interesse da aluna em tecnologia, informação na web e sistemas de recuperação de informação, que aliados às técnicas de tratamento e organização da informação, despertam a curiosidade e o interesse da aluna.

Com a popularização da Internet na sociedade moderna e o crescimento exponencial da web, torna-se fundamental a utilização de alternativas para a organização e recuperação das informações publicadas. Por ser um mecanismo que contribui para a organização e recuperação da informação, *Linked Open Data* pode ser usado como um aliado do profissional da informação no que se refere às publicações de dados abertos na web.

Em tempos em que boa parte da população tem acesso às Tecnologias da Informação e Comunicação, principalmente da Internet, é momento do profissional da informação rever seu importante papel, atuando de forma mais contundente na informação publicada na web e no tratamento e organização da informação também nos suportes digitais. Os dados publicados na web também devem ser objeto de estudo dos profissionais da informação e para isso é imprescindível que o profissional procure conhecer as ferramentas disponíveis para o tratamento da

informação digital. Sendo assim, pode-se dizer que o segundo motivo que justifica a pesquisa deve-se à necessidade de adequação do profissional da informação a este novo cenário.

Por último, o tema abordado justifica-se devido à escassez de estudos no que se refere à Ciência da Informação. É possível encontrar alguns trabalhos relacionados aos aspectos técnicos do *Linked Open Data* no âmbito da Ciência da Computação, por outro lado, no âmbito da Ciência da Informação e Biblioteconomia, esses estudos sofrem uma carência de publicações. Sendo assim, espera-se que este trabalho estimule de alguma forma os profissionais da informação para a produção de literatura neste sentido.

### 1.3 OBJETIVOS

Nesta seção, serão apresentados os objetivos norteadores para este estudo estando divididos em objetivo geral e os objetivos específicos.

#### 1.3.1 Objetivo Geral

Apresentar *Linked Open Data* como ambiente para publicação de dados abertos bibliográficos e de dados abertos da ciência na web, através da análise dos principais projetos existentes.

#### 1.3.2 Objetivos Específicos

Os objetivos específicos do trabalho são:

- a) Apresentar os Dados Abertos Ligados (*Linked Open Data*);
- b) Identificar a aplicação do *Linked Open Data* na publicação de dados bibliográficos e em dados da ciência e seus principais projetos;
- c) Apontar quais os tipos de informação que são disponibilizadas nos principais projetos e quais ontologias que estão sendo utilizadas para descrever essas informações;
- d) Constatar qual o estágio de desenvolvimento dos principais projetos;

- e) Identificar quais instituições que participam de cada um dos principais projetos;
- f) Verificar se as bases de dados dos projetos estão interligadas com outras bases de dados.

## 2 METODOLOGIA

Nesta seção, serão apresentados os princípios metodológicos utilizados para a realização do presente estudo.

### 2.1 ABORDAGEM E TIPO DE PESQUISA

A metodologia utilizada para realização da pesquisa foi a de abordagem qualitativa e de cunho exploratório. Segundo Gil (2010, p. 27), “As pesquisas exploratórias têm como propósito proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses. [...]”. Sendo assim, optou-se por adotar este tipo de pesquisa com a finalidade de se conhecer os conceitos e abordagens existentes sobre o tema e explorar as principais aplicações e projetos que utilizam o sistema.

### 2.2 CORPUS DA PESQUISA

O *corpus* do presente estudo é composto pelos projetos de instituições que estão utilizando o ambiente *Linked Open Data* para a disponibilização de seus dados na web. A amostra foi definida com base na revisão de literatura desenvolvida, sendo estabelecido o enfoque em dois tipos principais de aplicação: dados bibliográficos e dados da ciência. Os projetos selecionados para análise foram considerados a partir da relevância para o tema identificada durante o levantamento bibliográfico realizado. O quadro 1 apresenta os projetos que foram selecionados para análise e as respectivas instituições fundadoras de acordo com sua aplicação.

**Quadro 1 – Projetos Selecionados**

<b>FOCO: PUBLICAÇÃO DE DADOS BIBLIOGRÁFICOS</b>	
<b>PROJETO:</b>	<b>INSTITUIÇÃO:</b>
The British National Bibliography Linked Data	<ul style="list-style-type: none"> <li>• Biblioteca Britânica</li> </ul>
Virtual International Authority File (VIAF)	<ul style="list-style-type: none"> <li>• Online Computer Library Center (OCLC);</li> <li>• Biblioteca Nacional da Alemanha;</li> </ul>

	<ul style="list-style-type: none"> <li>• Biblioteca Nacional da França;</li> <li>• Biblioteca do Congresso Americano.</li> </ul>
Europeana Linked Data	<ul style="list-style-type: none"> <li>• Europeana</li> </ul>
<b>FOCO: PUBLICAÇÃO DE DADOS DA CIÊNCIA</b>	
<b>PROJETO:</b>	<b>INSTITUIÇÃO:</b>
Linked Open Data University of Münster (LODUM)	<ul style="list-style-type: none"> <li>• Universidade de Münster (Alemanha)</li> </ul>
Eagle-I	<ul style="list-style-type: none"> <li>• Dartmouth College;</li> <li>• Universidade de Harvard;</li> <li>• Universidade de Jackson State;</li> <li>• Morehouse School of Medicine (MSM);</li> <li>• Universidade do Estado de Montana;</li> <li>• Oregon Health &amp; Science University;</li> <li>• Universidade de Alaska Fairbanks;</li> <li>• Universidade do Havaí;</li> <li>• Universidade de Porto Rico.</li> </ul>

**Fonte:** Autora (2013).

## 2.3 INSTRUMENTOS DE COLETA DE DADOS

Para a coleta de dados do presente estudo foram utilizados dois instrumentos distintos. O primeiro instrumento consistiu na utilização de fichas (APÊNDICE A) contendo itens pré-estabelecidos para o controle dos dados coletados que serviram para o embasamento teórico da pesquisa. As fichas foram o instrumento utilizado para armazenar os principais conceitos extraídos das fontes de informação, transcritos com a respectiva referência bibliográfica e observações. O segundo instrumento utilizado para a coleta de dados foi um formulário de observação (APÊNDICE B), construído com itens pré-estabelecidos sobre os aspectos que se pretendeu observar nos projetos selecionados.

O quadro 2 estabelece as relações entre os objetivos específicos da pesquisa, os métodos utilizados e as fontes de onde foram extraídas as informações necessárias para alcançá-los.

**Quadro 2 – Objetivos e Coleta de Dados**

<b>Objetivo</b>	<b>Método</b>	<b>Fonte</b>
a) Apresentar os Dados Abertos Ligados ( <i>Linked Open Data</i> )	APÊNDICE A	Bem e Coelho (2013); Berners-Lee (1998); Berners-Lee (2006); Berners-Lee, Hendler e Lassila (2001); Brasil ([200-?]); Brickley ([200-?]); Brickley e Miller (2010); Feitosa (2006); Geonames ([201-?]); Hinz e Palazzo (2008); Lima e Carvalho (2005); Marcondes e Campos (2008); Mucheroni, Paiva e Netto (2009); Open Knowledge Foundation ([200-?], 2007); Rocha (2012); Santos Neto et al. (2013); Smith, Welty e McGuinness (2004); Souza e Alvarenga (2004); W3C (2001); Villalobos e Silva (2010).
b) Identificar a aplicação do <i>Linked Open Data</i> na publicação de dados bibliográficos e em dados da ciência e seus principais projetos	APÊNDICE A	Bourges-Waldegg (2013); Byrne e Goddard (2010); Kauppinen, Baglatzi e Kessler (2013); Kessler e Kauppinen (2012); Online Computer Library Center ([2012?]); Patrício (2012); Sandín Enríquez (2011); Santos Neto et al. (2013).
c) Apontar quais os tipos de informação que são disponibilizadas nos principais projetos e quais ontologias que estão sendo utilizadas para descrever essas informações	Questões 1 e 4 APÊNDICE B	British Library Research Explorer (2012); Eagle-I ([201-?]); Europeana ([201-?]); Linked Open Data University of Münster ([201-?]); Online Computer Library Center ([2012?]).
d) Constatar qual o estágio de desenvolvimento dos principais projetos	Questão 2 APÊNDICE B	British Library Research Explorer (2012); Eagle-I ([201-?]); Europeana ([201-?]); Linked Open Data University of Münster ([201-?]); Online Computer Library Center ([2012?]).
e) Identificar quais instituições que participam de cada um dos principais projetos	Questão 3 APÊNDICE B	British Library Research Explorer (2012); Eagle-I ([201-?]); Europeana ([201-?]); Linked Open Data University of Münster ([201-?]); Online Computer Library Center ([2012?]).
f) Verificar se as bases de dados dos projetos estão interligadas com outras bases de dados	Questão 5 APÊNDICE B	British Library Research Explorer (2012); Eagle-I ([201-?]); Europeana ([201-?]); Linked Open Data University of Münster ([201-?]); Online Computer Library Center ([2012?]).

**Fonte:** Autora (2013).

## 2.4 PROCEDIMENTOS DE COLETA DE DADOS

Quanto aos procedimentos, a pesquisa é bibliográfica, pois os dados foram coletados de material já publicado sobre o tema com o propósito de conhecer o que pensam os autores sobre o assunto. Segundo Pádua (2000, p. 52), a finalidade da pesquisa bibliográfica “[...] é colocar o pesquisador em contato com o que já se produziu e registrou a respeito do seu tema de pesquisa.” Este tipo de pesquisa permite que as reflexões feitas pelo pesquisador sobre o tema da pesquisa sejam fundamentadas em teorias e publicações de autores. Para Gil (2010, p. 30) “A principal vantagem da pesquisa bibliográfica reside no fato de permitir ao investigador a cobertura de uma gama de fenômenos muito mais ampla do que aquela que poderia pesquisar diretamente. [...]”.

A coleta de dados ocorreu em duas etapas. No primeiro momento, foi feito o levantamento bibliográfico, durante os meses de agosto e setembro de 2013, nas seguintes bases de dados: Brapci, LISA, SciELO e Google Acadêmico, bem como em catálogos de bibliotecas e periódicos. As principais expressões de busca utilizadas foram: “Linked Data”, “open data”, “dados abertos”, “dados ligados”; “Linked Science”, “Web Semântica”. Foram considerados somente os documentos publicados a partir do ano 2000, escritos no idioma português, espanhol ou inglês. Foram consideradas algumas publicações em *websites* devido à relevância indispensável para o desenvolvimento do trabalho. Após o levantamento da literatura, foi efetuada a leitura com apontamentos dos itens selecionados para extração dos dados. No segundo momento, foram selecionados os principais projetos identificados a partir da revisão de literatura desenvolvida e aplicou-se o formulário de pesquisa para a observação de cada um dos projetos. O instrumento foi aplicado em publicações específicas de cada um dos projetos e nas informações disponibilizadas nos *websites* dos projetos e das instituições vinculadas. Baseado nas informações coletadas a partir do preenchimento do formulário passou-se a fase de análise dos dados.

### 3 DADOS ABERTOS

Dados abertos (*open data*) são informações publicadas na web que podem ser acessadas e utilizadas livremente por qualquer pessoa com acesso à *Internet*. Segundo definição da Open Knowledge Foundation ([200-?], *online*) “Dados abertos são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa – sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras.” Para garantir o acesso livre e a plena utilização das informações, os dados devem ser publicados sem nenhum tipo de restrição de utilização e disponibilizados em formato que seja legível por máquina para que possam atingir o propósito de serem utilizados, reutilizados e distribuídos pelos usuários comuns.

A grande diferença entre dados fechados e dados abertos é que no primeiro tipo, as informações estão inseridas em uma base de dados e apenas usuários autorizados têm acesso a esses dados. Já os dados abertos são publicados na web livres de qualquer tipo de restrição de uso e qualquer pessoa pode ter acesso às informações.

A Internet é uma das principais fontes de informação da atualidade devido à sua adesão mundial e facilidade de utilização das ferramentas que a compõe. A disponibilização de dados para a população em geral tem a finalidade de permitir que qualquer pessoa ou organização possa dispor desses dados de forma bruta para operá-los da forma que lhe for conveniente, agregando e relacionando dados. Por exemplo, a publicação aberta de dados de um governo, pode permitir que qualquer cidadão atue na vigilância das ações deste governo analisando seus dados referentes aos gastos públicos deste governo. *Open data* contribui para a democratização dos fatos, ajudando a construir uma sociedade melhor informada, mais atuante e colaborativa.

Os dados abertos podem ser utilizados nas mais diferentes áreas e para as mais diversas aplicações. Algumas aplicações comuns são na área científica, bibliográfica e na área governamental. Na ciência, os dados abertos compreendem em informações provenientes de diferentes pesquisas científicas, que são publicadas na web de forma livre. O acesso a esses dados pode promover a colaboração entre os pesquisadores e a antecipação de descobertas. Entre os dados de ciência, estão os recursos que se referem aos pesquisadores, aos dados



institucionais e às pesquisas desenvolvidas. Na área bibliográfica, estão os dados referentes às publicações, autoridades, documentos e catálogos. São todas as informações extraídas de catálogos de bibliotecas e bases de dados bibliográficas. Na área governamental, dados abertos compreendem em dados referentes ao setor público. A publicação dos dados governamentais é fundamental para a promoção da transparência fiscal e vai ao encontro do direito de acesso a essas informações, que foi garantido ao cidadão através da Lei nº. 12.527, de 18 de novembro de 2011. Lei esta, que regula o acesso à informação e o artigo 5º, inciso XXXIII Constituição Federal.

A ideia de dados abertos governamentais surgiu como forma da sociedade cobrar informações públicas do governo. As informações do governo são aquelas correspondentes aos serviços públicos que dizem respeito à sociedade de uma forma geral, como despesas, receitas e informações de utilidade pública em geral. Conforme definição do Portal do Governo Eletrônico do Brasil:

Os Dados Abertos Governamentais são uma metodologia para a publicação de dados do governo em formatos reutilizáveis, visando o aumento da transparência e maior participação política por parte do cidadão, além de gerar diversas aplicações desenvolvidas colaborativamente pela sociedade. (BRASIL, [200-?], *online*).

A disponibilização dos dados do governo de forma livre traz efeitos benéficos para a sociedade, pois promove a maior participação do povo nas questões públicas e a democratização da informação pública. Os efeitos que a publicação de dados abertos governamentais pode trazer à sociedade são:

**Inclusão:** Fornecer dados em formatos padronizados abertos e acessíveis permite que qualquer cidadão utilize qualquer ferramenta de software para adaptá-los às suas necessidades.

**Transparência:** Informações do setor público abertas e acessíveis melhoram a transparência, pois as partes interessadas podem usá-las da maneira mais adequada ao seu propósito.

**Responsabilidade:** Os conjuntos apropriados de dados abertos, devidamente associados, podem oferecer vários pontos de vista sobre o desempenho do governo no cumprimento de suas metas em políticas públicas. (BRASIL, [200-?], *online*).

As iniciativas referentes à publicação de dados governamentais na web e as questões relativas aos efeitos da disponibilização destas informações à sociedade,

foram fatores que incentivaram a criação de regras básicas para que a publicação das informações de forma aberta na web aconteça de forma cada vez mais homogênea pelas instituições para possibilitar a utilização plena dos recursos que são disponibilizados.

Para que a publicação dos dados abertos governamentais aconteça de forma a garantir que os benefícios provenientes da sua disponibilização sejam alcançados, devem ser seguidos alguns princípios básicos. Sendo assim, em 2007, nos Estados Unidos, um grupo de trabalho promovido pela Open Knowledge Foundation estabeleceu os princípios para a publicação dos dados abertos governamentais na web, que hoje são utilizados com princípio para publicação de dados abertos em qualquer área:

- 1. Os dados devem estar completos:** [...] Dados públicos são dados que não estão sujeitos à privacidade, segurança ou privilégios e limitações válidas, que são regidas por outros estatutos.
- 2. Os dados devem ser primários:** Os dados são publicados como coletados na fonte, com o melhor nível possível de granularidade, não em formas agregadas ou modificadas.
- 3. Os dados devem ser atuais:** Os dados são disponibilizados com a rapidez necessária para que se mantenha seu valor.
- 4. Os dados devem estar acessíveis:** Os dados estão disponíveis para o maior número possível de usuários, para a mais ampla gama de propósitos.
- 5. Os dados devem ser processáveis por máquinas:** Os dados são razoavelmente estruturados para permitir um tratamento automatizado dos mesmos.
- 6. O acesso deve ser não discriminatório:** Os dados estão disponíveis para qualquer pessoa, sem necessidade de inscrição.
- 7. Dados devem ser de formato não proprietário:** Os dados estão disponíveis em um formato sobre o qual nenhuma entidade tem controle exclusivo.
- 8. Os dados devem ser livres de licenças:** Os dados não estão sujeitos a qualquer direito autoral, patente, marca registrada ou segredo comercial. Restrições de privacidade razoáveis, de segurança e privilégio podem ser permitidas, desde que regidas por outras leis. (OPEN KNOWLEDGE FOUNDATION, 2007, *online*, tradução nossa).

Algumas iniciativas nacionais e internacionais surgiram no sentido de promover a publicação, a utilização e o desenvolvimento de dados abertos. Tais iniciativas disponibilizam informações e recomendações para os órgãos públicos sobre a disponibilização dos dados abertos e para a sociedade sobre a utilização dos dados, bem como promover debates e trazer informações relativas ao impacto

social e os benefícios que estas publicações podem trazer à sociedade. Dentre estas, destacamos:

- a) **Open Knowledge Foundation:** é uma instituição de escala global, fundada em 2004, que tem por finalidade promover o conhecimento livre para toda a sociedade, através do incentivo à criação, disseminação e uso de conhecimento aberto, que pode ser utilizado, reutilizado e redistribuído sem restrições.
  
- b) **Consórcio World Wide Web (W3C):** é um consórcio internacional para o desenvolvimento de padrões para a web. É liderado pelo inventor da web Tim Berners-Lee e sua finalidade é garantir o crescimento da web de forma a atingir todo seu potencial. O W3C Brasil conta com quatro grupos de trabalho para lidar com as questões relacionadas à web: GT Certificação, GT Dados Abertos Governo e Sociedade, GT Vocabulários e Ontologias em Web Semântica e GT Acessibilidade Web.
  
- c) **Programa de Governo Eletrônico Brasileiro (GOV.BR):** é uma iniciativa governamental utilizada para democratizar o acesso à informação no país utilizando as tecnologias da informação e transformar a relação entre Governo e cidadãos. As principais finalidades do GOV.BR são promover uma maior participação cidadã e uma melhor administração pública. Através do projeto Portal Brasileiro de Dados Abertos são disponibilizadas a busca e acesso aos dados de transparência pública do Brasil.

O processo para a publicação de dados abertos na web necessita seguir metodologias estabelecidas pelas iniciativas que regulam as questões de disponibilização de dados abertos, para que a transparência pública aconteça de forma plena e as informações sejam acessadas e utilizadas de forma facilitada pelos cidadãos.

## 4 WEB SEMÂNTICA

*Linked Open Data* é uma forma de publicar dados de forma interligada, que utiliza como plataforma de publicação a Web Semântica. Este capítulo apresenta a Web Semântica, dando destaque às suas estruturas de representação (URI, RDF e OWL), utilizadas por *Linked Open Data* para a publicação de dados abertos, assim como ontologias utilizadas para publicar dados da ciência e bibliográficos.

A Web Semântica é a evolução da web atual, onde os dados são publicados semanticamente, utilizando ferramentas que dão estrutura e significados bem definidos aos recursos, permitindo que a informação seja inteligível tanto por máquinas como por pessoas. A idéia de Web Semântica surgiu em 2001, em um artigo de autoria de Berners-Lee, Hendler e Lassila publicado na revista *Scientific American*, que trouxeram a primeira definição de Web Semântica:

A Web Semântica não é uma Web separada, mas uma extensão da atual, em que é dado à informação um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação. Os primeiros passos para tecer a Web Semântica para a estrutura da Web existente já estão a caminho. Em um futuro próximo, estes desenvolvimentos vão inaugurar nova funcionalidade significativa como as máquinas se tornando muito mais capazes de processar e "entender" os dados que atualmente elas simplesmente apresentam. (BERNERS-LEE; HENDLER; LASSILA, 2001, p. 1, tradução nossa).

A transição da web atual para a Web Semântica pretende estabelecer contexto e conexão entre os dados, através da utilização de tecnologias e padrões recomendados para publicação dos recursos, buscando a evolução da web tradicional – a web dos documentos –, para uma nova web – a web dos dados –, em que os recursos sejam disponibilizados com conteúdo bem definido, de forma a garantir sua melhor interpretação e utilização, pelos usuários e pelas máquinas, conforme afirmam Souza e Alvarenga:

[...] A Web Semântica não trata de uma revolução, mas sim de uma evolução da Web como a conhecemos hoje. Trata-se principalmente da adoção de padrões de metadados e de compartilhamento destes padrões, de forma que possamos melhor utilizar o vasto repositório de informações disponível da Web de maneira mais produtiva, ágil e significativa. (SOUZA; ALVARENGA, 2004, p. 139).

A finalidade da Web Semântica é tornar a recuperação da informação na web mais eficiente através da atribuição de semântica às informações publicadas. Segundo Villalobos e Silva:

A Web Semântica tem uma finalidade clara, que é simplesmente melhorar a sistemática da recuperação da informação através de técnicas avançadas de representação da informação. Enquanto a Web atual utiliza um mecanismo de representação de caracteres voltado para o entendimento dos usuários, a Web Semântica promove a compreensão da informação através das máquinas, quer dizer, a Web Semântica terá a vantagem de estruturar os dados encontrados nos sites para que eles, antes de serem entendidos pelos usuários sejam entendidos pelo próprio sistema. (VILLALOBOS; SILVA 2010, p. 66).

Enquanto a web atual utiliza ferramentas que tornam possível a compreensão dos dados apenas pelos usuários, a Web Semântica estabelece padrões para atribuir semântica aos dados, tornando-os significativos também para máquinas e por conseqüência, tornam a recuperação da informação mais eficiente. São eles: *Uniform Resource Identifier* (URI) para a identificação única e persistente dos recursos, *Resource Description Framework* (RDF) como modelo básico de descrição semântica dos recursos, *Web Ontology Language* (OWL) como linguagem ontológica para definir e instanciar ontologias que serão utilizadas, e ontologias para representar os recursos, dando significado aos dados.

Os padrões estabelecidos pela iniciativa da Web Semântica para a publicação dos dados na web serão separadamente descritos nas subseções que seguem.

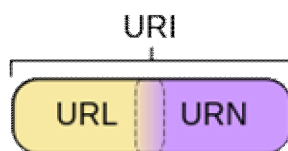
#### 4.1 UNIFORM RESOURCE IDENTIFIER (URI)

*Uniform Resource Identifier* (URI) são identificadores persistentes de recursos utilizados para referenciar dados e estabelecer ligações entre eles. A utilização de URIs para identificar recursos é requisito básico para a publicação de dados abertos vinculados. Rocha (2012, p. 283) afirma que “[...] Com o objetivo de permitir a ligação entre dados publicados por diversas organizações, *Linked Data* estabelece como princípio usar URIs para identificar unicamente as entidades representadas nesses conjuntos de dados.” Os dados são representados por URIs, que são *links* semânticos compreensíveis tanto pelas pessoas quanto pelas máquinas, que permitem a interoperabilidade dos recursos informacionais na web.

Conforme define Berners-Lee (1998, *online*, tradução nossa) “A Uniform Resource Identifier (URI) é um conjunto compacto de caracteres utilizado para identificar um recurso abstrato ou físico. [...]” Para o autor, a visão geral de URI pode ser caracterizada da seguinte forma: a **uniformidade** permite que diferentes identificadores de recursos sejam usados no mesmo contexto, mesmo quando os mecanismos utilizados para acessar esses recursos sejam diferentes. O **recurso** pode ser qualquer coisa que tenha identidade, como um documento eletrônico, uma imagem, um serviço, etc. O **identificador** é um objeto que funciona como referência para alguma coisa que tenha identidade. Quando um recurso é identificado, o sistema pode realizar operações como acessar, atualizar, substituir ou encontrar atributos.

O conceito de URI é muitas vezes confundido com os conceitos de *Uniform Resource Name* (URN) e *Uniform Resource Locator* (URL). Segundo W3C (2001, *online*, tradução nossa) “[...] URN é um esquema de URI, que define subespaços, chamados de “namespaces”. [...]” enquanto que “[...] a URL é um tipo de URI que identifica um recurso através de uma representação de seu principal mecanismo de acesso [...]”. Conforme ilustrado na Figura 1, podemos dizer que tanto URN como URL são subconjuntos do URI, sendo que o URN é utilizado para nomear os recursos e o URL é utilizado para indicar a localização dos recursos.

**Figura 1 – URI, URL e URN**



**Fonte:** Wikipédia (2013).

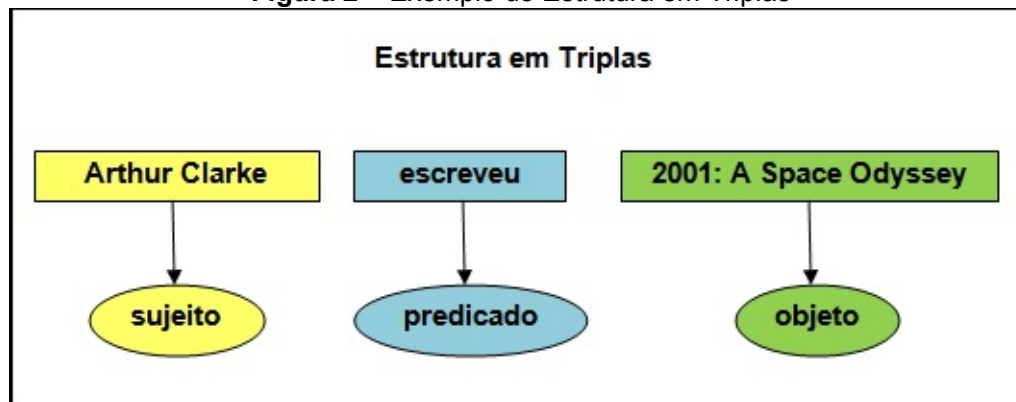
Segundo Feitosa (2006, p. 79) “[...] o uso de URIs permite que uma linguagem faça referência a objetos definidos em outra linguagem.” A utilização das URIs para identificação dos recursos possibilita que o item se torne único na web e sua identificação persistente, podendo ser associado a diversos outros recursos de forma interligada.

## 4.2 RESOURCE DESCRIPTION FRAMEWORK (RDF)

*Resource Description Framework* (RDF) é o modelo de dados básico para descrição semântica dos recursos da web e para publicação de *Linked Open Data*. No modelo RDF, os conteúdos são publicados de forma estruturada através de triplas, chamadas de sentenças, formadas por recurso (sujeito), predicado e objeto, que são identificados através de URIs. O sujeito pode ser qualquer recurso que possua uma identidade; o predicado é o que relaciona o sujeito ao objeto, e o objeto é aquilo que sofre influência do predicado, podendo também ser um recurso, conforme exemplificado na Figura 2.

Na figura 2 temos a sentença RDF “Arthur Clark escreveu 2001: A Space Odyssey”, onde **Arthur Clark** é o sujeito, **escreveu** é o predicado, e **2001: A Space Odyssey** é o objeto. Nessa sentença, **Arthur Clark** e **2001: A Space Odyssey** são representado por URIs, isto é, por endereços de páginas da web que representam, respectivamente o autor e o livro, que poderiam ser as páginas do autor e do livro na Wikipédia. A estrutura em triplas é o que possibilita a compreensão semântica dos recursos pelas máquinas.

**Figura 2 – Exemplo de Estrutura em Triplas**



Fonte: Autora (2013).

Segundo Feitosa (2006, p. 101) “[...] o padrão RDF foi projetado para permitir que *software* inteligente, como um *spider* ou um agente inteligente, consiga descobrir, catalogar e entender recursos em um *site* [...]”. O significado das triplas de RDF é representado através da *Web Ontology Language* (OWL). O modelo de dados RDF e a linguagem OWL são ferramentas utilizadas para representar semanticamente os recursos.

### 4.3 WEB ONTOLOGY LANGUAGE (OWL)

*Web Ontology Language* (OWL) é uma linguagem ontológica baseada no modelo de dados RDF utilizada para descrever semanticamente os recursos publicados na web. Segundo Smith, Welty e McGuinness (2004, *online*, tradução nossa) “A OWL Web Ontology Language é uma linguagem para definir e instanciar ontologias web. Ontologia é um termo emprestado da filosofia que se refere à ciência de descrever os tipos de entidades no mundo e como eles estão relacionados.”

As linguagens ontológicas são utilizadas na web para descrever os recursos e definir relações existentes entre eles e têm seus conceitos processáveis por máquinas. A linguagem OWL organiza os recursos a partir da descrição estruturada em classes, conforme afirma Feitosa (2006, p. 115):

Do mesmo modo que em RDF, a linguagem OWL utiliza estruturas de classes para a organização dos recursos. Assim, cada classe pode fornecer um mecanismo de abstração para o agrupamento de recursos e está associada a um conjunto de objetos individuais, a que se denomina extensão da classe. Cada objeto individual em uma classe é denominado instância de tal classe e possui um significado intencional. Devido a essa concepção, todos os recursos de um documento OWL são descritos como classes, subclasses ou propriedades de classes e subclasses. Isso garante, também, a linguagem, a possibilidade de representar significações com base em elementos que atuam como funções em algoritmos e que permitem a realização de inferências por agentes inteligentes. (FEITOSA, 2006, p. 115).

OWL consiste em uma linguagem padrão para ontologias estabelecido pelo W3C com o objetivo de contribuir para a evolução da web atual para a Web Semântica. O princípio fundamental da Web Semântica é a compreensão dos recursos tanto por máquinas como por pessoas e a utilização de linguagens ontológicas para representar o conteúdo de forma semântica é essencial para se alcançar tal compreensão.



#### 4.4 ONTOLOGIAS PARA WEB SEMÂNTICA

Ontologias são ferramentas utilizadas para representar o conhecimento e possibilitar que exista comunicação desse conhecimento tanto entre pessoas como entre máquinas. Para Lima e Carvalho (2005, p. 2) “Uma ontologia define um conjunto comum de termos que são usados para descrever e representar um domínio, como medicina, biblioteca etc. [...]”. A ontologia, segundo Mucheroni, Paiva e Netto (2009, p. 296) “[...] define os termos e suas relações em domínios, usando regras de combinação desses termos e relações.” O uso de ontologias objetiva o desenvolvimento da compreensão comum de um determinado conhecimento e seu compartilhamento. Para Hinz e Palazzo (2008, p. 1) “As Ontologias objetivam capturar conhecimento consensual entre pessoas e aplicações, permitindo o reuso e compartilhamento de informações. [...]”. Segundo Marcondes e Campos:

[...] Uma ontologia define um vocabulário comum para uma comunidade que precisa compartilhar informação em um determinado domínio. Inclui definições de conceitos básicos no domínio e as relações entre estes de forma que sejam interpretáveis por máquina. (MARCONDES; CAMPOS, 2008, p. 121).

A ontologia é uma ferramenta primordial para que a web atual evolua para a Web Semântica, pois através da utilização dos vocabulários e linguagens ontológicas para descrição de recursos na web, se agrega semântica às informações, tornando os recursos processáveis por humanos e por máquinas. Segundo Marcondes e Campos:

A proposta da Web Semântica prevê as ontologias como um dos pilares da sua arquitetura. As ontologias enquanto bases de conhecimento sobre domínios específicos estariam associadas a sistemas ou “web services”, guardando o conhecimento sobre e a semântica desses sistemas, de modo que programas inteligentes, os “agentes de software” pudessem interagir sem restrições com esses sistemas, tornando-os interoperáveis. (MARCONDES; CAMPOS, 2008, p. 117).

A utilização de ontologias permite que programas de computadores façam inferências, através das relações estabelecidas entre os termos, facilitando e enriquecendo a experiência de busca por informações na web. Lima e Carvalho afirmam que: “As ontologias podem ser usadas, dentre outras coisas, com o

propósito de melhorar a exatidão de buscas na Web: um programa de busca baseado em ontologias poderá ser capaz de recuperar somente as páginas relevantes para o usuário.” (LIMA; CARVALHO, 2005, p. 2). A utilização de ontologias está atrelada às questões de organização e recuperação da informação, conforme afirmam Bem e Coelho (2013):

De qualquer forma, sua relação está sempre associada à recuperação, armazenamento e uso da informação e conhecimento, dentro ou fora da web, seja para realizar tarefas, trocar e recuperar informações ou emular uma atividade humana, como no caso dos Sistemas Inteligentes. (BEM; COELHO, 2013, p. 155).

Entre as ontologias utilizadas para a descrição de recursos na web, destacam-se principalmente o Friend of a Friend (FOAF) e o GeoNames, que serão brevemente descritos.

#### **4.4.1 Friend of a Friend (FOAF)**

Friend of a Friend surgiu no início do ano 2000 e consiste em uma ontologia utilizada para descrever dados de pessoas, suas atividades e seus relacionamentos. Segundo definição de Hinz e Palazzo (2009, p. 4) “O projeto do FOAF tem por objetivo descrever as pessoas, as relações entre elas e as coisas que elas criam e fazem.” Para Brickley e Miller (2010):

[...] FOAF é um conjunto de definições concebidas para servir como um dicionário de termos que podem ser utilizados para expressar afirmações sobre o mundo. O foco inicial do FOAF foi a descrição de pessoas, pois as pessoas são as coisas que ligam a maioria dos outros tipos de coisas que descrevemos na Web: elas fazem documentos, participam de reuniões, são retratados em fotografias, e assim por diante. (BRICKLEY; MILLER, 2010, *online*, tradução nossa).

São objetivos do FOAF, segundo Brickley:

Queremos ser capazes de localizar documentos na Web com base nas suas propriedades e inter-relações, queremos ser capazes de encontrar informações sobre pessoas com base nas suas publicações, informações de emprego, participação em grupos e interesses declarados. Queremos ser capazes de compartilhar anotações, avaliações, favoritos e fragmentos de dados úteis

utilizando alguma infra-estrutura comum. Queremos um sistema de busca da Web que seja mais como um banco de dados e menos como um golpe de sorte. (BRICKLEY, [200-?], *online*, tradução nossa).

FOAF descreve os recursos de pessoas e suas relações de forma aberta e interligada e segue os requisitos estabelecidos pela W3C para a construção de uma web mais semântica, em que os recursos são inteligíveis por máquinas.

#### 4.4.2 GeoNames

GeoNames é uma ontologia utilizada para descrever dados geográficos através de uma base de dados que contém informações geográficas do mundo todo. São mais de 8 milhões de dados geográficos, entre topônimos em diferentes línguas, latitude, longitude, altitude, população e códigos postais. Os recursos são associados entre si e estão disponíveis para acesso e download no site oficial.

A ontologia GeoNames possibilita acrescentar informações semânticas geoespaciais para a World Wide Web. Todos os mais de 8,3 milhões topônimos de Geonames têm uma URL única, com um serviço de web RDF correspondente. Outros serviços descrevem a relação entre os topônimos. (GEONAMES, [201-?], *online*, tradução nossa).

A ontologia GeoNames contribui para uma web mais semântica, através da disponibilização de recursos de forma aberta e interligada e utilização das recomendações da W3C para publicação de dados ligados na web, como URI, RDF e OWL.

GeoNames é bastante utilizado e difundido em todo o mundo. Algumas das principais instituições que utilizam a ontologia GeoNames em seus sites e aplicativos são: Apple, Nokia, Bing Mapas, Foursquare, Europeana, The Time Now, Adidas, Nike, New York Times, BBC, The Guardian, entre outras.

## 5 LINKED OPEN DATA (LOD)

*Linked Open Data* (LOD) são dados abertos vinculados, ou seja, são dados abertos que estão interligados através de *links* externos e sua função é referenciar dados e documentos associados através de um sistema de *links* externos. Segundo Rocha (2012, p. 282-283) “*Linked Data* é o termo usado para descrever recomendações de melhores práticas para expor, compartilhar e conectar pedaços de dados e conhecimento na Web Semântica [...]”. *Linked Open Data* utiliza a Web Semântica como plataforma para a publicação dos dados abertos de forma estruturada e integrada através de *links* externos que referenciam dados relacionados.

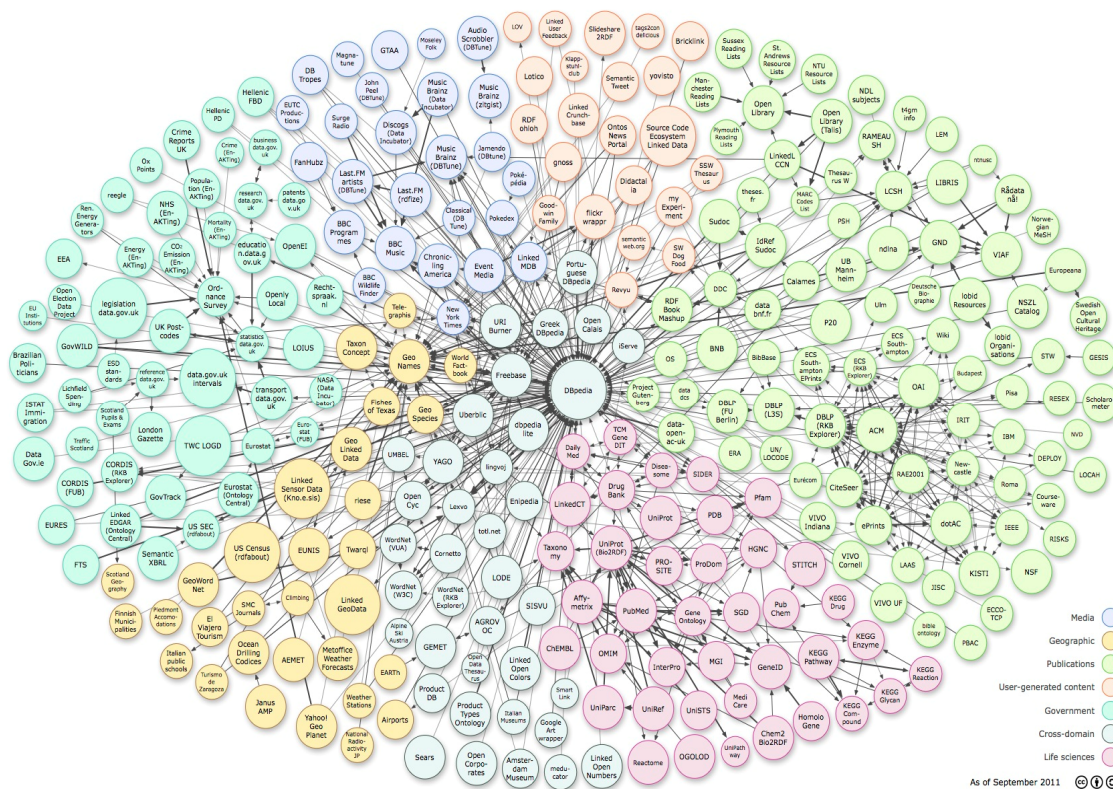
LOD é um método composto por recomendações que devem ser seguidas para a publicação de dados, com o objetivo de contribuir para uma web de dados interligados. Para Santos Neto e colaboradores (2013, p. 86) “O *Linked Data* aparece como uma possibilidade de facilitar a interligação de dados que, por sua vez, possam ser ligados a outros dados, criando uma rede de dados estruturados e com representação semântica legível por homens e máquinas.” Através de uma rede de dados representados por *links* vinculados é possível tornar a experiência do usuário mais enriquecedora, pois possibilita que as respostas para o problema inicial vinculem assuntos relacionados ao tema, permitindo que o conhecimento sobre o assunto ocorra de forma mais ampla e satisfatória.

A essência do *Linked Open Data* é a conexão das informações publicadas na web, criando contexto entre os dados e transformando assim, a web atual em uma web de dados, em que os recursos são todos interligados formando uma rede de informações relacionadas.

A utilização dos requisitos estabelecidos pelo *Linked Open Data* pelas instituições está crescendo de forma bastante significativa nos últimos anos. Instituições dos mais diferentes segmentos estão disponibilizando seus dados abertos de forma interligada a outras bases de dados, contribuindo para o crescimento da web dos dados, conforme pode ser visualizado nas figuras 3, 4 e 5, que apresentam o cenário do LOD nos anos 2007, 2009 e 2011, respectivamente. Estas figuras destacam as principais bases de dados no ambiente *Linked Open Data*, proposta de acordo com o estudo de Cyganiak e Jentzsch (2011),



Figura 5 – Diagrama de Nuvem dos Dados Abertos Ligados em 2011



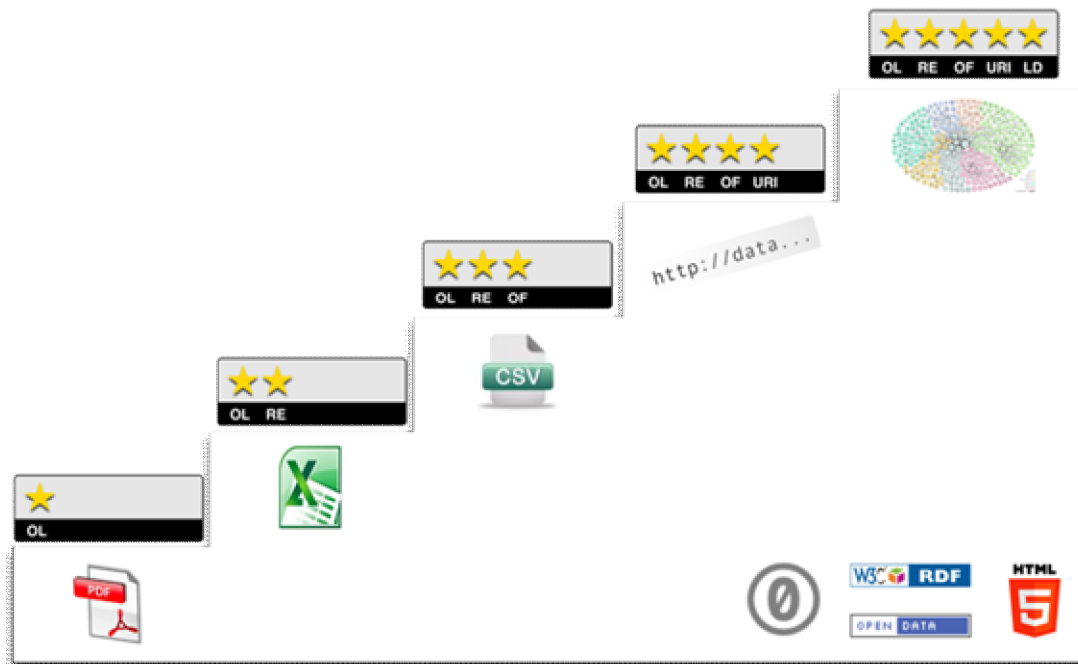
Fonte: Cyganiak e Jentzsch (2011).

Com a finalidade de avaliar o grau de reutilização e interoperabilidade dos dados abertos disponibilizados na web, Tim Berners-Lee (2006) criou um esquema baseado em estrelas. As estrelas indicam a forma com que os dados de uma base de dados estão sendo disponibilizados na web, sendo que bases de dados avaliadas com cinco estrelas são bases de dados publicadas como *Linked Open Data*:

- **1 Estrela:** Disponibilizar os dados na web com licença aberta.
- **2 Estrelas:** Disponibilizar os dados de forma estruturada (ex: uma planilha ao invés de uma imagem de uma tabela), além de atender os requisitos para uma estrela.
- **3 Estrelas:** Disponibilizar os dados utilizando formatos não-proprietários, além de atender os requisitos para duas estrelas.
- **4 Estrelas:** Utilizar os padrões estabelecidos pelo W3C: URI para identificar coisas e propriedades, além de atender os requisitos para três estrelas.
- **5 Estrelas:** Conectar seus dados com os dados de outras pessoas para criar contexto.

A figura 6 traz a representação gráfica do esquema baseado em estrelas proposto por Berners-Lee em 2006.

**Figura 6** – Esquema 5 estrelas para Open Data



**Fonte:** 5 Stars Open Data (2012).

**Legenda:** OL: Open License (licença aberta); RE: Machine-Readable (legível por máquina); OF: Open Format (formato aberto, não proprietário); URI: Uniform Resource Identifier (identificador uniforme de recurso); LD: Linked Data (dados ligados).

Um dos aspectos mais significativos da publicação de dados abertos de forma interligada é a utilização de *links* semânticos para a identificação de recursos. Os *links* semânticos são *links* significativos também para programas de computador, o que viabiliza que máquinas façam a associação de recursos com temas relacionados que são expressos por diferentes *links* semânticos.

Para que a publicação de dados abertos na web ocorra segundo as recomendações do *Linked Open Data*, de modo a tornar possível a interligação entre os recursos publicados, atingindo assim seu objetivo de transformar a web convencional em uma web de dados relacionados, é fundamental seguir os requisitos necessários para a publicação dos dados.

Com a finalidade de usufruir ao máximo os recursos oferecidos pela web, Berners-Lee (2006) indica quatro princípios que devem ser seguidos para publicação de dados abertos na web:

- a) Usar **Uniform Resource Identifier** (URI) para identificar os dados;
- b) Utilizar o protocolo **Hypertext Transfer Protocol** (http) para facilitar a localização dos dados;
- c) Fornecer informações úteis e utilizar padrões como o **Resource Description Framework** (RDF);
- d) Incluir **links** para outras URIs para que os usuários possam descobrir mais informações relacionadas.

Não seguir os princípios anteriormente citados, segundo Berners-Lee (2006), não impede que os dados sejam publicados, mas se perde a oportunidade de tornar os dados interligados e possibilitar sua reutilização.

Devido aos princípios necessários para a publicação dos dados abertos utilizando *Linked Open Data*, a plataforma ideal para a disponibilização dos dados é a Web Semântica, pois possibilita grande interação e recuperação da informação através de seus recursos para identificação, representação e interligação de recursos publicados. Segundo Rocha (2012, p. 290):

[...] A plataforma da Web Semântica proporciona vantagens para a integração e interligação de dados. O uso de URI como identificador evita redundância uma vez que um dado externo pode ser referenciado, ao invés de ser localmente replicado. A integração da informação é facilitada, pois ocorre no nível semântico e não no nível estrutural, visto que RDF e OWL são instrumentos para adicionar semântica aos recursos, sem fazer suposição sobre a estrutura desses recursos. (ROCHA, 2012, p. 290).

A publicação de dados na web deve ocorrer através de *links* semânticos que, ao contrário de *links* convencionais, também são significativos para máquinas, que conseguem compreender a semântica dos recursos e fazer associações com outros *links* externos. Para que a publicação de dados na web ocorra utilizando *links* semânticos é essencial a utilização de alguns padrões específicos da Web Semântica, como *Uniform Resource Identifier* (URI) para a identificação dos recursos, *Resource Description Framework* (RDF) para a representação dos recursos e *Web Ontology Language* (OWL) para a descrição semântica dos recursos. A utilização de tais requisitos para a publicação de dados na web é fundamental para que ocorra a interação entre os dados.



*Linked Open Data* é um mecanismo que agrega contexto às informações, pois através da interligação dos dados, promove o desenvolvimento contínuo da exploração de conhecimento em diferentes abordagens.

Diferentes tipos de instituições e pessoas podem se beneficiar da utilização das ferramentas de dados abertos vinculados. Para a presente pesquisa, o enfoque do estudo foi *Linked Open Data* na publicação de dados bibliográficos, provenientes de disponibilização de recursos de catálogos de bibliotecas e bases de dados bibliográficos, e dados da ciência, provenientes de recursos extraídos de ativos da ciência.

## 5.1 BIBLIOTECAS LIGANDO DADOS

Alguns catálogos de bibliotecas estão acessíveis para o usuário na Internet através dos sistemas específicos utilizados por cada instituição. Porém, de maneira geral, esses sistemas, embora utilizem esquemas padrões de representação, como Marc e Dublin Core, não permitem que os seus dados sejam interligados com dados de catálogos de bibliotecas de instituições distintas. Segundo Byrne e Goddard (2010):

Ter um formato comum para todos os dados seria um enorme benefício para a interoperabilidade e a integração de todos os tipos de sistemas. Pode-se facilmente imaginar bibliotecas que trabalham com seus fornecedores para desenvolver colaborativamente uma grande base de conhecimento compartilhado [...]. (BYRNE; GODDARD, 2010, *online*, tradução nossa).

A integração de dados pode auxiliar as bibliotecas nas diferentes tarefas necessárias para seu funcionamento como “[...] as funções de seleção, ordenação, catalogação, controle de autoridade, desenvolvimento de taxonomia e pesquisa. [...]” (BYRNE; GODDARD, 2010, *online*, tradução nossa) através do trabalho colaborativo, utilização, reutilização e disponibilização dos dados. Para os autores:

Linked Data representa uma grande oportunidade para as bibliotecas integrarem seus recursos de informação com a web em geral. Além disso, fornece uma base para uma melhor gestão de dados, permitindo bibliotecas armazenarem, compartilharem e reutilizarem dados, conforme necessário. A tecnologia está finalmente pronta, é fundamental para as bibliotecas iniciar os preparativos para tornarem-se participantes plenas do mundo dos dados vinculados. (BYRNE; GODDARD, 2010, *online*, tradução nossa).

A utilização de dados de bibliotecas abertos e ligados promove a interoperabilidade das informações, redução de incoerências e enriquecimento informacional para as entidades participantes. Sobre as iniciativas de bibliotecas ligando dados, Santos Neto e colaboradores afirmam que:

Já existem diversas iniciativas de dados abertos interligados. Um exemplo é a Biblioteca Europeia, desenvolvida pela União Europeia. Nela, pode-se encontrar diversas representações de documentos e imagens de obras de arte provenientes de diferentes instituições de diversos países europeus, interligadas entre si. Outras iniciativas são provenientes da Biblioteca do Congresso dos Estados Unidos, da British Library, da Biblioteca Nacional da Alemanha, da Biblioteca do Centro George Pompidou na França e da Biblioteca Nacional da Espanha. (SANTOS NETO et al., 2013, p. 82).

Atualmente, existem grandes projetos de bibliotecas que estão publicando seus catálogos de forma aberta e interligada a outras bases de dados, utilizando os princípios do *Linked Open Data*. Entre os principais projetos, destacam-se os seguintes: The British National Bibliography (BNB) Linked Data, o Virtual International Authority File (VIAF) e o Europeana Linked Data.

### **5.1.1 Projetos de Dados Bibliográficos em Linked Open Data**

Foram analisados 3 projetos de instituições que utilizam o ambiente *Linked Open Data* para publicação de dados bibliográficos: The British National Bibliography (BNB) Linked Data, Virtual International Authority File (VIAF) e Europeana Linked Data. Os resultados da análise desses projetos serão descritos a seguir.

#### **5.1.1.1 The British National Bibliography (BNB) Linked Data**

A Biblioteca Britânica (*British Library*) registra as atividades de publicação do Reino Unido e da República da Irlanda, desde 1950, funcionando como uma medida da produção intelectual dos países. Tradicionalmente, os registros eram de publicações impressas e mais recentemente foi prorrogado para as publicações eletrônicas. Isso ocorreu a partir de 2003, quando o depósito legal foi estendido para este tipo de material.

*The British National Bibliography* ou a Bibliografia Nacional Britânica é um projeto da Biblioteca Britânica que consiste na aplicação das recomendações do *Linked Open Data* para publicação dos seus registos bibliográficos.

[...] Os benefícios de uma abordagem de *Linked Open Data* para a Biblioteca Britânica incluem: melhor integração web de seus recursos com maior visibilidade e oportunidades para alcançar novos usuários, maior flexibilidade com novas opções de ligação para um rico conjunto global de dados reutilizáveis para melhorar os serviços e reduzir esforço duplicado, novas perspectivas, competências da equipe e contatos de colaboração e de melhores opções para os desenvolvedores que desejam trabalhar com os metadados da Biblioteca Britânica. (BRITISH, 2012, *online*, tradução nossa).

A Bibliografia Nacional Britânica disponibiliza os metadados dos registos bibliográficos de livros e publicações seriadas publicados no Reino Unido de forma aberta e ligada, seguindo o modelo de dados RDF, a linguagem ontológica OWL e as seguintes ontologias compatíveis com a Web Semântica:

- Dublin Core: metadados básicos para descrever recursos.
- FOAF: para descrever as pessoas e suas relações.
- GeoNames: para descrição de recursos geoespaciais.
- Bibliographic Ontology (BIBO): para descrição de publicações e seus recursos relacionados.
- Bio – A Vocabulary for Biographical Information: vocabulário para publicação de informações biográficas.
- British Library Terms: vocabulário para complementar esquemas bibliográficos existentes.
- Event Ontology: termos para descrever os eventos e seus participantes.
- International Standard Bibliographic Description (ISBD): para descrição bibliográfica dos diferentes tipos de documentos.
- Org: An Organization Ontology: termos de apoio para publicação de dados abertos ligados organizacionais.
- Simple Knowledge Organization System (SKOS): termos de apoio para a publicação de classificações de assunto e taxonomias.
- WGS84 Geo Positioning: para a representação informações geográficas.

- Versão preliminary do Resource Description and Access (RDA): para descrição de recursos bibliográficos.

O projeto está em desenvolvimento contínuo e atualmente, o número de registros disponibilizados é estimado em aproximadamente 3 milhões de registros. Sua base de dados está ligada com outras bases de dados, o que permite a interoperabilidade e a cooperação entre as entidades vinculadas. As bases de dados ligadas à BNB Linked Data são:

- Virtual International Authority File (VIAF): base de dados de autoridades da Online Computer Library Center (OCLC).
- Library of Congress Subject Headings (LCSH): base de dados de cabeçalhos de assuntos da Library of Congress.
- Lexvo: base de dados sobre idiomas, palavras e outras entidades lingüísticas na web.
- GeoNames: base de dados geográficos.
- MARC Country and Language Codes: base de dados de países, línguas e grupos lingüísticos.
- Dewey.info: base de dados das notações da Classificação Decimal de Dewey (CDD).
- RDF Book Mashup: base de dados de informações sobre os livros, seus autores, comentários e livrarias online disponíveis na web.

O quadro 3 apresenta versão resumida das informações coletadas na análise do projeto BNB Linked Data.

**Quadro 3 – The British National Bibliography (BNB) Linked Data**

<b>The British National Bibliography (BNB) Linked Data</b>	
Tipo de informação:	Metadados de registros de livros e publicações seriadas
Estágio de desenvolvimento:	<ul style="list-style-type: none"> <li>• Desenvolvimento contínuo</li> <li>• Cerca de 3 milhões de dados</li> </ul>
Instituições participantes:	Biblioteca Britânica
Ontologias / padrões utilizados:	<ul style="list-style-type: none"> <li>• Dublin Core</li> <li>• FOAF</li> </ul>

	<ul style="list-style-type: none"> <li>• GeoNames</li> <li>• BIBO</li> <li>• Bio</li> <li>• British Library Terms</li> <li>• Event Ontology</li> <li>• ISBD</li> <li>• Org</li> <li>• SKOS</li> <li>• WGS84 Geo Positioning</li> <li>• Versão preliminar do RDA</li> </ul>
Bases de dados ligadas:	<ul style="list-style-type: none"> <li>• VIAF</li> <li>• LCSH</li> <li>• Lexvo</li> <li>• GeoNames</li> <li>• MARC Country and Language Codes</li> <li>• Dewey.info</li> <li>• RDF Book Mashup</li> </ul>

**Fonte:** Autora (2013).

#### 5.1.1.2 Virtual International Authority File (VIAF)

A Base Virtual Internacional de Autoridade (Virtual International Authority File – VIAF) é um serviço de arquivos de autoridades virtual e internacional formado pela interligação dos registros de autoridades das bibliotecas participantes. Segundo definição da Online Computer Library Center ([2012?], online):

O Catálogo de autoridade internacional virtual (VIAF) é um serviço internacional criado para facilitar o acesso aos maiores catálogos de autoridade de nomes do mundo. Seus criadores o idealizaram como uma base para a Web semântica que permitisse a alteração do modo de exibição dos nomes de pessoas, para que eles pudessem ser exibidos com o idioma e script escolhidos pelo usuário da Web. (ONLINE COMPUTER LIBRARY CENTER, [2012?], *online*).

O projeto teve início em 2003 com a iniciativa da Online Computer Library Center (OCLC) em parceria com a Biblioteca Nacional da Alemanha (Deutsche

Nationalbibliothek), Biblioteca Nacional da França (Bibliothèque National de France) e Biblioteca do Congresso Americano (Library of Congress).

Segundo Sandín Enríquez (2011, p. 24, tradução nossa), o objetivo principal do VIAF é “[...] introduzir os registros de autoridades de diferentes bibliotecas, em um único serviço disponível na rede. [...]” Para o autor, o serviço permite que os usuários de diferentes lugares do mundo pesquisem uma autoridade em uma base de dados única e os resultados são apresentados em diferentes línguas. Sendo assim, os objetivos específicos do VIAF são:

- Relacionar os registros de autoridades, tanto em nível nacional como regional.
- Divulgar a noção de controle bibliográfico universal através da união de formas autorizadas de diferentes países, assim como, proporcionar diferentes preferências de idioma, escrita e ortografia.
- Tentar ser um elemento importante para a Web Semântica. (SANDÍN ENRÍQUEZ, 2011, p. 24, tradução nossa).

Segundo a OCLC, o VIAF pode ser resumido como:

- Um esforço colaborativo entre bibliotecas nacionais e organizações contribuindo com arquivos de autoridade de nome, aumentando o acesso a informações.
- Todos os dados de autoridade para uma dada entidade estão vinculados em outro "super" registro de autoridade.
- Uma maneira conveniente para a comunidade de bibliotecas e outras agências reutilizarem dados bibliográficos produzidos por bibliotecas atuantes em diferentes comunidades de idiomas. (ONLINE COMPUTER LIBRARY CENTER, [2012?], *online*).

Atualmente, o VIAF possui mais de 20 milhões de registros e envolve a colaboração de um número crescente de parceiros. Segundo dados da OCLC, em 2012, o número de parceiros chegou a 20 instituições em 16 países (ONLINE COMPUTER LIBRARY CENTER, [2012?]).

Para descrever seus recursos, o projeto VIAF as seguintes ontologias:

- GeoNames: para descrever dados geográficos.
- FOAF: para descrever as pessoas e suas relações.
- Dublin Core: metadados básicos para descrever recursos.

- SKOS: termos de apoio para a publicação de classificações de assunto e taxonomias.

Virtual International Authority File interliga sua base de dados outras bases externas, garantindo a interoperabilidade do serviço. As bases de dados ligadas ao VIAF são:

- WorldCat Identities, da OCLC: base de dados dos registros de coleções das bibliotecas participantes da Online Computer Library Center.
- GeoNames: base de dados geográficos.

Por se tratar de uma cooperação internacional, apesar do registro de cada autoridade ser único, a apresentação dos dados é feita em diferentes idiomas, conforme descreve a Online Computer Library Center ([2012?], *online*): “[...] Os usuários de cada país também podem ver os registros de nomes no formato estabelecido por entidades de outros países, o que torna as autoridades realmente internacionais e facilita as pesquisas em qualquer idioma.”

A aplicação dos padrões do Linked Open Data em catálogos, como o serviço oferecido pelo VIAF, possibilita a interoperabilidade e a cooperação entre as entidades participantes, aspectos que contribuem para a redução de custos das bibliotecas participantes, conforme apresentado pela OCLC:

O VIAF ajuda a reduzir os custos de manutenção dos catálogos de autoridade e a torná-los mais úteis para as bibliotecas e usuários em geral. Para isso, o VIAF compara e associa os catálogos de autoridade das bibliotecas nacionais e agrupa todos os registros de autoridade de uma determinada entidade em um “super” registro de autoridade que contém os diferentes nomes fornecidos para essa entidade. Ao associar nomes distintos fornecidos a uma mesma pessoa ou organização, o VIAF permite que um número maior de bibliotecas e outras agências reutilizem convenientemente dados bibliográficos produzidos por bibliotecas que atendem a comunidades com diferentes idiomas. (Online Computer Library Center, [2012?], *online*).

O VIAF contribui para uma web mais semântica, pois utiliza ontologias e interliga as informações de diferentes bibliotecas. Além disso, o serviço cria um registro único para cada autoridade e permite o suporte aos diferentes idiomas e

ortografias de acordo com as especificações do usuário, o que evidencia a universalidade e interoperabilidade do VIAF.

O quadro 4 apresenta uma versão resumida das informações coletadas na análise do projeto VIAF.

**Quadro 4 – Virtual International Authority File (VIAF)**

<b>Virtual International Authority File (VIAF)</b>	
Tipo de informação:	Metadados de registros de autoridades.
Estágio de desenvolvimento:	<ul style="list-style-type: none"> <li>• Iniciou em 2003.</li> <li>• Está em desenvolvimento contínuo com colaboração crescente de colaboradores.</li> <li>• Cerca de 20 milhões de dados</li> </ul>
Instituições participantes:	20 instituições em 16 países.
Ontologias / padrões utilizados:	<ul style="list-style-type: none"> <li>• GeoNames</li> <li>• FOAF</li> <li>• Dublin Core</li> <li>• SKOS</li> </ul>
Bases de dados ligadas:	<ul style="list-style-type: none"> <li>• WorldCat Identities da OCLC</li> <li>• GeoNames</li> </ul>

**Fonte:** Autora (2013).

### 5.1.1.3 Europeana Linked Data

A Europeana é uma grande biblioteca virtual que recolhe dados de objetos digitais de galerias, bibliotecas, arquivos e museus de toda a Europa com a finalidade de proporcionar um grande conjunto de dados culturais europeus.

O projeto piloto da Europeana Linked Data surgiu em 2012 com a abertura e interligação dos registros da Europeana. “A versão atual do piloto contém metadados de 20 milhões de textos, imagens, vídeos e sons recolhidos pela Europeana.” (EUROPEANA, [201-?], *online*, tradução nossa). São metadados extraídos de itens de mais de 1500 instituições culturais europeias integrantes da Europeana.

A adoção de uma arquitetura LOD para a Europeana visa facilitar a agregação de dados de diferentes setores culturais, o enriquecimento dos dados agregados com recursos de contextualização disponíveis na web e a reutilização de dados da Europeana como recursos de contextualização para outros recursos



da web. No caso específico dos dados bibliográficos, a sua disponibilização na web semântica implica decisões ao nível da formalização de metadados que só agora começam a ser equacionadas e que, a par de alterações já em curso, tornarão os dados criados pelas bibliotecas e os recursos por eles descritos mais visíveis e reutilizáveis fora do seu contexto original. [...] (PATRÍCIO, 2012, p. 11-12).

O Europeana Linked Data utiliza as seguintes ontologias e padrões de amplo uso para a descrição dos dados:

- Dublin Core: ontologia para metadados básicos.
- FOAF: para descrever as pessoas e suas relações.
- CIDOC Conceptual Reference Model (CRM): ontologia para descrição de recursos de documentação, património e museus.
- Open Archives Initiative Object Reuse and Exchange (OAI-ORE): padrão para descrição de recursos na web.
- SKOS: termos de apoio para a publicação de classificações de assunto e taxonomias.

Quanto à interoperabilidade, a base de dados da Europeana Linked Data liga sua base de dados com outras bases, permitindo a colaboração entre os serviços oferecidos. As bases de dados ligadas são:

- DBpedia: base de dados estruturados extraídos da Wikipédia.
- VIAF: base de dados de autoridades.

O quadro 5 apresenta uma versão resumida das informações coletadas na análise do projeto Europeana Linked Data.

**Quadro 5 – Europeana Linked Data**

<b>Europeana Linked Data</b>	
Tipo de informação:	Metadados de registros de objetos digitais de galerias, bibliotecas, arquivos e museus
Estágio de desenvolvimento:	<ul style="list-style-type: none"> <li>• Surgiu em 2012 e está em desenvolvimento contínuo</li> <li>• Cerca de 20 milhões de dados</li> </ul>

Instituições participantes:	Cerca de 1.500 instituições culturais de toda a Europa
Ontologias / padrões utilizados:	<ul style="list-style-type: none"> <li>• Dublin Core</li> <li>• FOAF</li> <li>• OAI-ORE</li> <li>• SKOS</li> <li>• CIDOC CRM</li> </ul>
Bases de dados ligadas:	<ul style="list-style-type: none"> <li>• DBpedia</li> <li>• VIAF</li> </ul>

**Fonte:** Autora (2013).

## 5.2 A CIÊNCIA LIGANDO DADOS

*Linked Science* é o termo utilizado para denominar os dados provenientes da ciência publicados de forma conectada. É uma iniciativa para interligar ativos da ciência, como dados e resultados de pesquisas, informações sobre pesquisadores e publicações, entre outros, através dos requisitos estabelecidos pelo *Linked Open Data*, conforme definem Kauppinen, Baglatzi e Kessler (2013, p. 16, tradução nossa) *Linked Science* é “[...] uma abordagem onde recursos científicos – processos, modelos, dados, métodos e métricas de avaliação – estão semanticamente ligados.”

Ligar ativos da ciência facilita a produção de novas pesquisas, pois a disponibilização dos dados abertos e ligados auxilia na promoção do trabalho colaborativo entre os pesquisadores, permitindo a reutilização dos dados e principalmente a transparência nas pesquisas desenvolvidas, conforme apontam Kessler e Kauppinen (2012, p. 1, tradução nossa) “Os diferentes recursos científicos, sejam eles publicações, conjuntos de dados, métodos ou ferramentas devem ser anotados, interligados e compartilhados abertamente, a fim de fazer a ciência mais transparente e reprodutível.”

Um dos grandes benefícios da utilização do *Linked Open Data* para a publicação dos dados da ciência está relacionado à redução de espaço de armazenamento de informações relevantes à pesquisa ou ao pesquisador, conforme apontam Kauppinen, Baglatzi e Kessler:

[...] a distribuição de dados através da Web reduz o espaço necessário para armazenar dados em um ambiente único e local. Por exemplo, se um cientista utilizar uma informação geográfica – lugares com suas coordenadas e limites poligonais e assim por

diante – ele apenas conecta para bases de dados já existentes, fornecendo esta informação ao invés de baixar tudo em seu ambiente local. (KAUPPINEN; BAGLATZI; KESSLER, 2013, p. 8, tradução nossa).

Segundo Kauppinen, Baglatzi e Kessler (2013, p. 15, tradução nossa) a distribuição dos dados da ciência de forma aberta e ligada “[...] procura ajudar os autores, revisores, editores e toda a comunidade científica em suas tarefas desafiadoras. [...]”. Ocorrendo através da facilitação do acesso às informações, da transparência nas investigações e da possibilidade da reutilização e redistribuição dos dados, proporcionadas pela estrutura na publicação dos dados.

Atualmente, diversas instituições estão publicando seus ativos de ciência utilizando os requisitos estabelecidos pelo *Linked Open Data*. Dois projetos se destacam na publicação do *Linked Science*: *Linked Open Data University of Münster* (LODUM) e *Eagle-I*.

#### **4.2.1 Projetos de Dados da Ciência em Linked Open Data**

Foram analisados 2 projetos de instituições que utilizam o ambiente *Linked Open Data* para publicação de dados da ciência: *Linked Open Data University of Münster* (LODUM) e *Eagle-I*. Os resultados da análise desses projetos serão descritos a seguir.

##### **5.2.1.1 Linked Open Data University of Münster (LODUM)**

O LODUM é um projeto da University of Münster, na Alemanha, que consiste em uma estratégia de acesso aberto para as publicações da instituição, seguindo os princípios do *Linked Open Data*. A estratégia é considerada de longo prazo, que vai abrir e conectar diferentes fontes de dados através das 15 faculdades e departamentos da Universidade, conforme apresentado na página do projeto:

LODUM é definido como uma estratégia de longo prazo que vai abrir e conectar diferentes fontes de dados através das 15 faculdades e departamentos passo a passo. [...] um aspecto crucial do LODUM é promover uma mudança na abertura e transparência, da Universidade de Münster e sua comunidade científica. Com a implementação do acesso aberto e princípios do *Linked Open Data* por toda a universidade, LODUM visa tornar a produção da

universidade mais visível e promover a colaboração, tanto entre as faculdades da universidade, como com os parceiros. (LINKED OPEN DATA UNIVERSITY OF MÜNSTER, [201-?], *online*, tradução nossa).

O projeto LODUM surgiu como a “[...] implementação de uma estratégia que visa melhorar a transparência e visibilidade da universidade.” (LINKED OPEN DATA UNIVERSITY OF MÜNSTER, [201-?], *online*, tradução nossa). O foco do projeto é a disponibilização de todo conteúdo relacionado ao que é produzido na instituição, como dados científicos e publicações, informações sobre pessoas, projetos, patentes e os prêmios da universidade.

O projeto tem como finalidade a promoção de mudanças na abertura e transparência das informações da Universidade de Münster e sua comunidade científica, além da colaboração entre as faculdades da universidade e com os parceiros do projeto. Para Kessler e Kauppinen (2012), preservar e facilitar o acesso dos recursos científicos é fundamental para garantir a transparência dos estudos. Sendo assim, o alicerce do LODUM é a preservação, armazenamento e acesso dos dados científicos e assuntos relacionados à pesquisa, além das questões administrativas da instituição.

LODUM é um trabalho colaborativo com a biblioteca da Universidade, coordenado pelo Laboratório de Interoperabilidade Semântica e gabinete do reitor da universidade.

O projeto LODUM utiliza ontologias já existentes e de amplo uso para descrever seus recursos. As ontologias utilizadas são:

- BIBO: para recursos bibliográficos como livros e artigos.
- Dublin Core: para metadados básicos.
- FOAF: para descrição de pessoas e suas relações.
- W3C Basic Geo Vocabulary: para descrever áreas geográficas.
- Teaching Core Vocabulary (TEACH): para descrição dos recursos e relações específicos dos professores e seus cursos.
- Publishing Requirements for Industry Standard Metadata (PRISM): para descrição de propriedades adicionais de metadados.

O quadro 6 apresenta uma versão resumida das informações coletadas na análise do projeto Linked Open Data University of Münster.

**Quadro 6 – Linked Open Data University of Münster (LODUM)**

<b>Linked Open Data University of Münster (LODUM)</b>	
Tipo de informação:	Metadados de registros científicos e publicações, informações sobre pessoas, projetos, patentes e os prêmios da universidade
Estágio de desenvolvimento:	<ul style="list-style-type: none"> <li>• Desenvolvimento contínuo</li> </ul>
Instituições participantes:	15 faculdades e departamentos da Universidade de Münster
Ontologias / padrões utilizados:	<ul style="list-style-type: none"> <li>• Dublin Core</li> <li>• BIBO</li> <li>• FOAF</li> <li>• W3C Basic Geo Vocabulary</li> <li>• TEACH</li> <li>• PRISM</li> </ul>

**Fonte:** Autora (2013).

#### 5.2.1.2 Eagle-I

O projeto Eagle-I surgiu em 2011, como um consórcio de instituições acadêmicas, com a finalidade de facilitar a publicação, o acesso e a utilização de recursos da área biomédica das universidades participantes do consórcio. As nove instituições acadêmicas que fundaram e deram início aos trabalhos desenvolvidos pelo Eagle-i são:

- Dartmouth College;
- Universidade de Harvard;
- Jackson State University;
- Morehouse School of Medicine (MSM);
- Montana State University;
- Oregon Health & Science University;
- Universidade de Alaska Fairbanks;
- Universidade do Haváí;
- Universidade de Porto Rico.

Segundo definição do *website* do projeto, “Eagle-i é um aplicativo gratuito, baseado na web que torna mais fácil descobrir os recursos disponíveis nas universidades participantes, ajudando os investigadores a acelerar suas pesquisas e fazer mais com menos.” (EAGLE-I, [201-?], *online*, tradução nossa). A iniciativa para o desenvolvimento do projeto se deve principalmente a dificuldade encontrada pelos pesquisadores na busca e no acesso aos dados provenientes de pesquisas da área biomédica, conforme descreve Bourges-Waldegg (2013, *online*, tradução nossa):

#### **Motivação**

Incapacidade dos cientistas para identificar e acessar os recursos críticos de investigação – seja localmente ou em outras instituições – é um grande obstáculo para o avanço da ciência e da medicina. O desenvolvimento de uma infra-estrutura nacional permitiria aos pesquisadores descobrir recursos vitais de pesquisa, acelerando, assim, o ritmo de tradução e melhorando a nossa capacidade de desenvolver novos diagnósticos, tratamentos e estratégias de prevenção.

O serviço pretende contribuir para a interoperabilidade e colaboração entre as universidades, para que se tornem parceiras na publicação e reutilização dos dados provenientes das pesquisas biomédicas, publicados de forma aberta e interligada. E para isso, utiliza os padrões estabelecidos pelo *Linked Open Data*, conforme afirma Bourges-Waldegg (2013, *online*, tradução nossa) o “Eagle-i é uma plataforma distribuída para a criação e compartilhamento de dados semanticamente ricos. Ele é construído em torno de tecnologias de web semântica e segue os princípios do Linked Open Data. [...]”.

Eagle-i utiliza para a descrição de seus recursos algumas ontologias de amplo uso na web e uma ontologia própria. As ontologias utilizadas são:

- Ontology for Biomedical Investigation (OBI): para descrição de recursos de investigações biológicas e clínicas;
- VIVO: para descrição dos recursos de pesquisadores em relação aos seus interesses, experiências, resultados e instituições relacionadas;
- Dublin Core: metadados básicos para descrever recursos.
- FOAF: para descrever as pessoas e suas relações.
- Uber-anatomy Ontology (Uberon): para descrição dos fenótipos de diferentes espécies;

- Ontology for Clinical Research (OCRe): para descrição recursos de estudos em seres humanos;
- Sequence Ontology (SO): para descrever seqüências biológicas;
- Software Ontology (SWO): para descrição de ferramentas de *software*.
- NCBI Taxonomy: para descrição de espécies;
- Eagle-I Research Resource Ontology (ERO): ontologia própria utilizada para descrever protocolos, reagentes, animais e bioespécimes.

Atualmente, vinte e seis instituições acadêmicas fazem parte da rede Eagle-i e mais entidades estão em processo de adesão. O projeto conta com mais de 54 mil recursos da área biomédica, sendo que este número aumenta consideravelmente a cada semana.

O quadro 7 apresenta uma versão resumida das informações coletadas na análise do projeto Eagle-I.

**Quadro 7 – Eagle-I**

<b>Eagle-I</b>	
Tipo de informação:	Metadados de recursos da área biomédica
Estágio de desenvolvimento:	<ul style="list-style-type: none"> <li>• Desenvolvimento contínuo, com adesão contínua de novas instituições</li> <li>• Cerca de 54 mil recursos</li> </ul>
Instituições participantes:	26 instituições
Ontologias / padrões utilizados:	<ul style="list-style-type: none"> <li>• OBI</li> <li>• VIVO</li> <li>• Dublin Core</li> <li>• FOAF</li> <li>• Uberon</li> <li>• OCRe</li> <li>• SO</li> <li>• SWO</li> <li>• NCBI Taxonomy</li> <li>• ERO</li> </ul>

**Fonte:** Autora (2013).

### 5.3 ONTOLOGIAS E BASES DE DADOS

Dos projetos analisados nesta pesquisa, todos utilizam como ontologias para a descrição de seus recursos: Dublin Core – como metadados básicos para a descrição dos recursos – e FOAF – para a descrição de pessoas e suas relações.

A ontologia SKOS é utilizada nos três projetos de *Linked Open Data* em dados bibliográficos para as classificações de assuntos e taxonomias.

GeoNames é utilizado como ontologia para descrever recursos geográficos de dois, dos três projetos analisados referentes aos dados abertos bibliográficos, BNB Linked Data e VIAF.

BIBO é a ontologia utilizada para descrição de recursos bibliográficos em dois projetos: BNB Linked Data (dados bibliográficos) e LODUM (dados da ciência). Outras ontologias são utilizadas para descrever os recursos específicos de cada um dos projetos analisados.

Os projetos analisados compartilham as ontologias mais gerais e amplamente conhecidas na web, como Dublin Core, FOAF, SKOS e GeoNames. Isso demonstra que existe uma preocupação por parte desses projetos na reutilização de ontologias, proporcionando com isso uma maior integração semântica entre os projetos que utilizam o ambiente *Linked Open Data* para publicação de seus dados na web.

O quadro 8 apresenta a relação entre os projetos analisados e as ontologias utilizadas, demonstrando a frequência do uso de determinadas ontologias pelos projetos analisados.

**Quadro 8 – Comparativo das Ontologias utilizadas nos Projetos**

ONTOLOGIA	PROJETO				
	BNB	VIAF	Europeana	LODUM	Eagle-I
BIBO	x			x	
Bio	x				
British Library Terms	x				
CRM			x		
Dublin Core	x	x	x	x	X
ERO					X
Event Ontology	x				
FOAF	x	x	x	x	X



<b>GeoNames</b>	<b>x</b>	<b>x</b>			
<b>ISBD</b>	<b>x</b>				
<b>NCBI Taxonomy</b>					<b>X</b>
<b>OAI-ORE</b>			<b>x</b>		
<b>OBI</b>					<b>X</b>
<b>OCRe</b>					<b>X</b>
<b>Org</b>	<b>x</b>				
<b>PRISM</b>				<b>x</b>	
<b>SKOS</b>	<b>x</b>	<b>x</b>	<b>x</b>		
<b>SO</b>					<b>X</b>
<b>SWO</b>					<b>X</b>
<b>TEACH</b>				<b>x</b>	
<b>Uberon</b>					<b>X</b>
<b>Versão preliminar do RDA</b>	<b>x</b>				
<b>VIVO</b>					<b>X</b>
<b>W3C Basic Geo Vocabulary</b>				<b>x</b>	
<b>WGS84 Geo Positioning</b>	<b>x</b>				

Fonte: Autora (2013).

Outro aspecto analisado nos projetos que utilizam *Linked Open Data* para publicar seus dados foi se as bases de dados destes projetos estavam interligadas com outras bases de dados, possibilitando a interoperabilidades entre as bases.

Na análise dos projetos de *Linked Open Data* para a publicação de dados bibliográficos observou-se que o VIAF possui sua base de dados interligada com as bases dos outros projetos analisados, demonstrando que existe uma preocupação dos projetos com o controle de autoridade. Tanto BNB Linked Data, quanto a Europeana Linked Data buscam associar suas autoridades ao VIAF.

A base de dados geográficos GeoNames e a base de dados de coleções de bibliotecas WorldCat Identities também estão conectadas com os projetos do VIAF e do BNB Linked Data.

Outras bases de dados estão conectadas às bases dos projetos que foram analisados. O quadro 9 apresenta a relação entre os projetos analisados e as base de dados externas que eles estão interligados, demonstrando a frequência que as bases são ligadas pelos projetos.

**Quadro 9 – Comparativo das Bases de Dados ligadas aos Projetos**

BASE DE DADOS	PROJETO		
	BNB	VIAF	Europeana
VIAF	x	x	X
LCSH	x		
Lexvo	x		
GeoNames	x	x	
MARC Country and Language Codes	x		
Dewey.info	x		
RDF Book Mashup	x		
WorldCat Identities	x	x	
DBpedia			X

**Fonte:** Autora (2013).

Na análise dos projetos de *Linked Open Data* para a publicação de dados da ciência, não foram encontradas informações sobre bases externas que estão ligadas às bases dos projetos. Apenas, constatou-se a conexão com os dados das diferentes instituições participantes de cada projeto e a possibilidade de se fazer *download* dos recursos publicados em cada um dos projetos.

## 6 CONSIDERAÇÕES FINAIS

A pesquisa desenvolvida teve como objetivo apresentar o *Linked Open Data* como ambiente para a publicação de dados abertos bibliográficos e dados abertos da ciência na web, através da análise dos principais projetos existentes.

Para se alcançar tal objetivo, buscou-se em um primeiro momento, fazer o levantamento bibliográfico de assuntos que pudessem conceitualizar e contextualizar *Linked Open Data* com a finalidade de garantir bagagem teórica para que a etapa seguinte do estudo pudesse ser desenvolvida.

No segundo momento, efetuou-se a pesquisa propriamente dita, com a finalidade de se conhecer os principais projetos desenvolvidos na utilização de LOD para a publicação dos dados abertos bibliográficos e dos dados abertos da ciência. A análise dos projetos teve como finalidade descobrir quais são os padrões e ferramentas utilizadas em cada área e os benefícios da utilização do ambiente LOD para a publicação dos dados bibliográficos e da ciência na web.

Na fase de levantamento bibliográfico, constatou-se que não existem muitas publicações sobre a temática no âmbito da Ciência da Informação, sendo o assunto mais estudado pela Ciência da Computação. Dos poucos estudos encontrados sob a ótica da Ciência da Informação, a grande maioria eram pesquisas desenvolvidas fora do Brasil. A partir deste cenário, conclui-se que ainda não há um interesse dos pesquisadores da Ciência da Informação, principalmente no Brasil, nas pesquisas relacionadas com *Linked Open Data*.

A análise dos projetos que utilizam LOD para a publicação de dados abertos bibliográficos e dados abertos da ciência, demonstrou a viabilidade e os benefícios da utilização desse ambiente para a publicação dos dados e as ferramentas que são mais utilizadas no cenário *Linked Open Data*.

Após a análise dos projetos concluiu-se que as ontologias mais utilizadas para a descrição dos recursos em LOD são: FOAF (descrição de informações sobre as pessoas), Dublin Core (metadados básicos para a descrição dos recursos), SKOS (descrição de classificações), GeoNames (descrição de recursos geográficos) e o BIBO (descrição de recursos bibliográficos). A análise dos projetos também possibilitou que se conhecessem as principais bases de dados externas que estão interligadas às bases de dados dos projetos que utilizam LOD: a base de

autoridades VIAF, a base de dados geográficos GeoNames e a base de dados de coleções de bibliotecas WorldCat Identities.

Conclui-se que os projetos analisados buscam interoperabilidade semântica, visto que utilizam as mesmas ontologias, principalmente as básicas, como FOAF, Dublin Core e GeoNames, e que suas bases de dados são ligadas entre si, como por exemplo, o VIAF, que está ligado às bases da BNB Linked Data e da Europeana Linked Data. Analisando os projetos, verificamos que eles estão engajados no que se refere à interoperabilidade, pois compartilham ontologias e ligam as mesmas bases de dados.

A interoperabilidade é o principal motivo para a publicação de dados abertos e ligados. Utilizar os padrões de interoperabilidade recomendados pela iniciativa da Web Semântica e *Linked Open Data* possibilita que os dados sejam utilizados, reutilizados, analisados e disponibilizados de forma a facilitar o uso das informações pelas instituições, pelos usuários e pelos profissionais da informação.

As instituições podem se beneficiar da utilização do ambiente *Linked Open Data* no que se refere à agregação de valor nos serviços prestados, já que podem contar com a utilização de dados de outras empresas sobre o mesmo cliente ou mercado, por exemplo. O trabalho colaborativo das instituições e a reutilização de informações são aspectos que podem ajudar a economizar recursos materiais e humanos para a instituição, ocasionando melhorias nos aspectos financeiros.

Para o usuário da web, o ambiente LOD melhora a experiência na busca por informações, pois a publicação de dados de forma estruturada e a utilização de ontologias proporcionam que as máquinas entendam melhor o que aquela busca quer dizer e por conseqüência, a pesquisa se torna menos penosa e o resultado mais adequado ao que o usuário necessita.

Para o profissional da Ciência da Informação, *Linked Open Data* é uma área de atuação em potencial que pode ser explorada. A tendência é que a web atual evolua para a Web Semântica e cada vez mais as instituições publiquem seus dados de forma aberta e ligada. O profissional da informação deve trabalhar para melhorar a organização e recuperação da informação na web, de forma que a disseminação da informação para a sociedade ocorra da melhor forma possível.

As Bibliotecas e centros de informação só ganham com a publicação de seus catálogos e recursos utilizando o ambiente *Linked Open Data*. A conexão entre coleções, acervos e catálogos permite o trabalho colaborativo entre as instituições,

interoperabilidade entre os recursos e agrega valor aos dados publicados. Essas questões trazem economia de recursos para a instituição, melhoram os serviços prestados, agregam contexto às informações disponibilizadas e enriquecem a experiência do usuário.

*Linked Open Data* como ambiente para a publicação de dados da ciência na web pode trazer muitos benefícios para os pesquisadores, para a Ciência e para a sociedade. A disponibilização dos dados provenientes de ativos da ciência de forma aberta permite que outros pesquisadores utilizem esses dados para fazer novos estudos e descobertas. Esse trabalho colaborativo promove a transparência dos resultados e a agilidade na execução de pesquisas pelos cientistas. Por consequência, a Ciência e a sociedade ganham com a maior rapidez, divulgação e transparência nas descobertas dos pesquisadores.

A pesquisa desenvolvida demonstrou que o *Linked Open Data* é um ambiente interessante para a publicação de dados abertos bibliográficos e dados abertos da ciência na web, pois segue a tendência de evolução da web atual para a Web Semântica e é uma forma de organização da informação na web que atribui semântica aos recursos, tornando-os compreensíveis também por máquinas e que, por consequência, melhora a recuperação da informação na web.

Existem projetos envolvendo universidades e bibliotecas nacionais importantes e esses projetos estão em estado avançado de desenvolvimento, não podendo ser considerados simples experimentos. São ambientes desenvolvidos principalmente para o progresso da ciência e da web. Isso sinaliza uma tendência favorável ao uso e a pesquisa de *Linked Open Data* na ciência e em bibliotecas. A partir deste cenário, em que grandes instituições e universidades estão aderindo ao LOD, pode-se concluir que investir nesse ambiente para publicação de dados na web é realmente vantajoso para o avanço da ciência e para a evolução da organização da informação na web.

A partir do estudo desenvolvido, pode-se concluir que LOD é um ambiente promissor para publicação de dados abertos bibliográficos e de dados abertos da ciência na web, e que pesquisas e projetos neste sentido, devem ser incentivados. Os profissionais da informação têm um papel importante a ser desempenhado no que se refere à utilização de LOD como ambiente para a publicação de dados na web, pois devem trabalhar por uma organização e recuperação da informação mais eficiente, independente de seu suporte.

## REFERENCIAS BIBLIOGRÁFICAS

5 STARS Open Data. [S.l.: s.n.], 2012. Disponível em: <<http://5stardata.info/>>. Acesso em: 19 out. 2013.

BEM, Roberta Moraes; COELHO, Christianne Coelho de Souza Reinisch. Instrumentos de Representação do Conhecimento para Práticas de Gestão do Conhecimento: taxonomias, tesouros e ontologias. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 4, n. 1, p. 147-162, jan./jun. 2013. Disponível em: <<http://www.revistas.usp.br/incid/article/view/59106>>. Acesso em: 31 out. 2013.

BERNERS-LEE, Tim. **Linked data**: design issues. 2006. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 12 jun. 2013.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, may 2001.

BERNERS-LEE, Tim et al. **Uniform Resource Identifiers (URI)**: Generic Syntax. 1998. Disponível em: <<http://www.ietf.org/rfc/rfc2396.txt>>. Acesso em: 17 jun. 2013.

BOURGES-WALDEGG, Daniela. **Welcome to eagle-i**. Cambridge: Harvard College, 2013. Disponível em: <<https://open.med.harvard.edu/display/eaglei/Welcome+to+eagle-i>>. Acesso em: 22 out. 2013.

BRASIL. Ministério do Planejamento, Orçamento e Gestão. **Dados Abertos Governamentais**. Portal do Governo Eletrônico do Brasil. [200-?]. Disponível em: <<http://www.governoeletronico.gov.br/acoes-e-projetos/Dados-Abertos>>. Acesso em: 24 maio 2013.

BRICKLEY, Dan. **Introducing FOAF**. [S.l.: s.n.], [200-?]. Disponível em: <<http://www.foaf-project.org/original-intro>>. Acesso em: 22 out. 2013.

BRICKLEY, Dan; MILLER, Libby. **FOAF Vocabulary Specification**. [S.l.: s.n.], 2010. Disponível em: <<http://xmains.com/foaf/spec/>>. Acesso em: 22 out. 2013.

BRITISH Library Research Explorer. [S.l.]: DATA.GOV.UK, 2012. Disponível em: <<http://data.gov.uk/library/british-library-research-explorer>>. Acesso em: 22 out. 2013.

BYRNE, Gillian; GODDARD, Lisa. The Strongest Link: libraries and Linked Data. **D-Lib Magazine**, v. 16, n. 11/12, nov./dec. 2010. Disponível em: <<http://www.dlib.org/dlib/november10/byrne/11byrne.html>>. Acesso em: 03 set. 2013.

CYGANIAK, Richard; JENTZSCH, Anja. **Linking Open Data Cloud Diagram**. Berlin: Freie Universitat, 2011. Disponível em: <<http://lod-cloud.net/>>. Acesso em: 16 out. 2013.

EAGLE-I. **What is eagle-i?** Cambridge: Harvard College, [201-?]. Disponível em: <<http://www.eagle-i.net/about/what-is-eagle-i/>>. Acesso em: 22 out. 2013.

EUROPEANA. **Linked Open Data** – data.europeana.eu. The Hague: Europeana, [201-?].

FEITOSA, Ailton. **Organização da Informação na Web: das Tags à Web Semântica**. Brasília, DF: Thesaurus, 2006.

GEONAMES. **GeoNames Ontology**. [S.l., s.n.]: [201-?]. Disponível em: <<http://www.geonames.org/ontology/documentation.html>>. Acesso em: 02 nov. 2013.

GIL, Antonio Carlos. **Como elaborar Projetos de Pesquisa**. 5. ed. São Paulo, SP: Atlas, 2010.

HINZ, Verlani Timm; PALAZZO, Luiz Antônio Moro. Algoritmos para Interoperabilidade entre Ontologias. In: SEMINÁRIO DE PESQUISA EM ONTOLOGIA NO BRASIL, 2008, Niterói. **Anais...** Niterói: Universidade Federal Fluminense, 2008. Disponível em: <<http://www.uff.br/ontologia/artigos/211.pdf>>. Acesso em: 20 out. 2013.

HINZ, Verlani Timm; PALAZZO, Luiz Antônio Moro. Interoperabilidade em Ambientes Educacionais Virtuais Modelados por Ontologias. In: CONGRESSO NACIONAL DE AMBIENTES HIPERMÍDIA PARA APRENDIZAGEM, 4., 2009, Florianópolis. **Anais...** [S.l.: s.n.], 2009. Disponível em: <[http://wright.ava.ufsc.br/~alice/conahpa/anais/2009/cd\\_conahpa2009/papers/final153.pdf](http://wright.ava.ufsc.br/~alice/conahpa/anais/2009/cd_conahpa2009/papers/final153.pdf)>. Acesso em: 20 out. 2013.

KAUPPINEN, Tomi; BAGLATZI, Alkyoni; KESSLER, Carsten. Linked Science: Interconnecting Scientific Assets. In: CRITCHLOW, Terence; DAM, Kerstin Kleese Van. (Eds.). **Data Intensive Science**. New York: CRC Press, 2013.

KESSLER, Carsten; KAUPPINEN, Tomi. Linked Open Data University of Münster – Infrastructure and Applications. In: EXTENDED SEMANTIC WEB CONFERENCE, 9., Heraklion, 2012. **Proceedings...** Heraklion: [s.n.], 2012. Disponível em: <[http://2012.eswc-conferences.org/sites/default/files/eswc2012\\_submission\\_333.pdf](http://2012.eswc-conferences.org/sites/default/files/eswc2012_submission_333.pdf)>. Acesso em: 25 set. 2013.

LIMA, Júnio César de; CARVALHO, Cedric Luiz de. **Ontologias - OWL (Web Ontology Language)**. 2005. Disponível em: <[http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_004-05.pdf](http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_004-05.pdf)>. Acesso em: 17 jun. 2013.

LINKED OPEN DATA UNIVERSITY OF MÜNSTER. **About LODUM**. Münster: University of Münster, [201-?]. Disponível em: <<http://lodum.de/about/>>. Acesso em: 25 out. 2013.

MARCONDES, Carlos Henrique; CAMPOS, Maria Luiza de Almeida. Ontologia e Web Semântica: o espaço da pesquisa em Ciência da Informação. **Ponto de Acesso**, Salvador, v. 2, n. 1, p. 107-136, jun./jul. 2008. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000005073&dd1=e283d>>. Acesso em: 31 out. 2013.

MUCHERONI, Marcos Luiz; PAIVA, Daniel C. de; NETTO, Marcio Lobbo. Três Ontologias Clássicas e a Web Semântica. **Ponto de Acesso**, Salvador, v. 3, n. 2, p. 281-298, dez. 2009. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000008222&dd1=e9761>>. Acesso em: 31 out. 2013

ONLINE COMPUTER LIBRARY CENTER. **VIAF**: catálogo de autoridade internacional virtual. Dublin: OCLC, [2012?]. Disponível em: <<http://www.oclc.org/pt-americalatina/viaf.html>>. Acesso em: 22 out. 2013.

OPEN KNOWLEDGE FOUNDATION. **Open Data Handbook**. [200-?]. Disponível em: <[http://opendatahandbook.org/pt\\_BR/what-is-open-data/index.html](http://opendatahandbook.org/pt_BR/what-is-open-data/index.html)>. Acesso em: 24 maio 2013.

OPEN KNOWLEDGE FOUNDATION. **8 Principles of Open Government Data**. 2007. Disponível em: <<http://www.opengovdata.org/home/8principles>>. Acesso em: 25 jun. 2013.

PÁDUA, Elisabete Matallo Marchesini de. **Metodologia da Pesquisa**: abordagem teórico-prática. 6. ed. Campinas, SP: Papyrus, 2000.

PATRÍCIO, Helena Simões. A Europeana e a agregação de metadados na web: análise dos esquemas ESE/EDM e da aplicação de standards da web semântica a



dados de bibliotecas. **Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas**, n. 11, 2012. Disponível em: <<http://www.bad.pt/publicacoes/index.php/congressosbad/article/view/458>>. Acesso em: 22 out. 2013.

ROCHA, Rafael Port da. Fabrico/Ciência: um ambiente Linked Data para o mapeamento da Ciência. **Em Questão**, Porto Alegre, v. 18, Edição Especial, p. 281-297, dez. 2012.

SANDÍN ENRÍQUEZ, Diego. **El Fichero Internacional Virtual de Autoridades (VIAF): estado de la cuestión, evaluación y propuestas de mejora**. 2011. 96 f. Dissertação (Mestrado em Sistemas de Informação Digital)–Facultad de Traducción y Documentación, Universidad de Salamanca, Salamanca, 2011. Disponível em: <<http://gredos.usal.es/jspui/handle/10366/116279>>. Acesso em: 20 set. 2013.

SANTOS NETO, Antonio Laurindo dos. et. al. Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. **TransInformação**, Campinas, v. 25, n. 1, p. 81-87, jan./abr. 2013. Disponível em: <<http://www.scielo.br/pdf/tinf/v25n1/a08v25n1.pdf>>. Acesso em: 11 jun. 2013.

SMITH, Michael K.; WELTY, Chris; MCGUINNESS, Deborah L. **OWL Web Ontology Language Guide**. 2004. Disponível em: <<http://www.w3.org/TR/2004/REC-owl-guide-20040210/>>. Acesso em: 17 jun. 2013.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para a Ciência da Informação. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004.

VILLALOBOS, Ana Paula de Oliveira; SILVA, Daniel Cerqueira. As potencialidades da Web Semântica para a Ciência da Informação. **Ponto de Acesso**, Salvador, v. 4, n. 2, p. 58-75, set. 2010. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/3623>>. Acesso em: 30 out. 2013.

W3C. **URIs, URLs, and URNs: Clarifications and Recommendations 1.0**. 2001. Disponível em: <<http://www.w3.org/TR/uri-clarification/>>. Acesso em: 17 jun. 2013.

WIKIPEDIA. **Uniform Resource Identifier**. [S.l.: s.n.], 2013. Disponível em: <[https://en.wikipedia.org/wiki/Uniform\\_resource\\_identifier](https://en.wikipedia.org/wiki/Uniform_resource_identifier)>. Acesso em: 19 jun. 2013.

**APÊNDICE A – Fichamento dos Dados**

<b>Assunto:</b>		
Referência Bibliográfica:		
Fonte:		
Data da Análise: ____/____/____		
<b>Página:</b>	<b>Citação:</b>	<b>Observações:</b>
<b>Observações Gerais:</b>		

## APÊNDICE B – Formulário de Coleta de Dados

<b>PROJETO:</b>	
<i>Website:</i>	
Data da Análise: ____/____/____	
<b>INFORMAÇÕES GERAIS SOBRE O PROJETO</b>	
1) Quais os tipos de informação utilizados (recursos)?	
2) Qual o início e o estágio de desenvolvimento do projeto?	
<b>UNIVERSALIDADE E ABRANGÊNCIA:</b>	
3) O projeto envolve qual (is) instituição (s)?	
4) O projeto usa quais ontologias de amplo uso?	
5) A base de dados está “linkada” com outras bases de dados? Quais?	
Observações	
Fontes Consultadas:	